DE GRUYTER
OPEN

Sentencing Disparities

Melissa Hamilton*
University of Surrey School of Law

ABSTRACT

Criminal justice stakeholders are strongly concerned with disparities in penalty outcomes. Disparities are problematic when they represent unfounded differences in sentences, an abuse of discretion, and/or potential discrimination based on sociodemographic characteristics. The Article presents an original empirical study that explores disparities in sentences at two levels: the individual case level and the regional level. More specifically, the study investigates upward departures in the United States' federal sentencing system, which constitutes a guidelines-based structure. Upward departures carry unique consequences to individuals and their effects on the system as they lead to lengthier sentences, symbolically represent a dispute with the guidelines advice, and contribute to mass incarceration. Upward departures are discretionary to district courts and thus may lead to disparities in sentencing in which otherwise seemingly like offenders receive dissimilar sentences, in part because of the tendency of their assigned judges to depart upward (or not).

The study utilizes a multilevel mixed model to test the effects of a host of explanatory factors on the issuance of upward departures at the case level and whether those same factors are significant at the group level-i.e., district courts-to determine the extent of variation across districts. The explanatory variables tested include legal factors (e.g., final offense level, criminal history, offense type), extralegal characteristics (e.g., gender, race/ethnicity, citizenship), and case-processing variables (e.g., trial penalty, custody status). The results indicate that various legal and nonlegal factors are relevant in individual cases (representing individual differences) and signify that significant variations across district courts exist (confirming regional disparities). Implications of the significant findings for the justice system are explored.

CONTENTS

---

*   Senior Lecturer in Law & Criminal Justice, University of Surrey School of Law; J.D., The University of Texas at Austin School of Law; Ph.D, The University of Texas at Austin (Criminology/Criminal Justice). The author expresses appreciation for helpful comments on an earlier draft from Benjamin Levin, J.C. Oleson, Anna Roberts and Justin Murray.

I. Introduction

Policymakers, practitioners, and academics have long brought attention to unjustified variations in criminal justice outcomes.[1] A principal focus is on disparities in sentencing practices because of the perception that inconsistencies in penalties are indicative of disproportionality in penalty outcomes, an abuse of discretion, and potential discrimination.[2] An additional concern today is America's evolution into a state of mass incarceration with too many individuals being sent to prison and for longer periods of time.[3] To investigate the possible existence of disparities, researchers from diverse academic disciplines have undertaken a host of studies.[4]

Nevertheless, there is much still to be learned. Serious gaps exist in the empirical legal studies literature regarding certain sentencing practices. The modal approaches to sentencing research is to focus on the in/out decision (i.e., whether the penalty requires any time of imprisonment) and sentence length.[5] Yet, there are other types of sentencing decisions that deserve more attention as they may also substantively exacerbate disparities in outcomes while contributing to mass incarceration. Then, more sophisticated empirical methodologies are available today that permit researchers to better specify statistical models to improve fit to the data and reduce the potential for biases in the results. Plus, there is perhaps insufficient attention to regional variations in sentencing practices.

This Article contributes to the literature by producing an empirical study focusing on sentences that constitute upward departures from sentencing guidelines. In particular, federal sentencing is a guidelines-based system, with upward departures issued at the discretion of district judges. Decisions to depart upward are uniquely remarkable because they obviously lead to lengthier prison terms, may represent gaps in the guidelines, and may signify disparities—potentially discrimination—in sentencing decisions. The federal system is worthy of analysis as it often acts as a role model for criminal justice practices, it operates the largest prison system in the country in terms of the number of inmates held, and it represents sentencing decisions across the country.

To date, no research appears to have discretely concentrated on upward departure decisions in federal sentencing. The results presented herein are meant to address this void. This study takes advantage of multilevel modeling as the empirical methodology, which constitutes a more sophisticated model of statistical analysis than is used in most criminal justice research.[6] The study also responds to a call

---

[1]   Michael Tonry, Sentencing Matters 4 (1996).

[2]   Cassia Spohn, *Twentieth-Century Sentencing Reform Movement: Looking Backward, Moving Forward*, 13 Criminology & Pub. Pol'y 535, 537 (2014).

[3]   Craig Haney, Reforming Punishment: Psychological Limits to the Pains of Imprisonment 61 (2006).

[4]   *See generally* Leslie Sebba, *Is Sentencing Reform a Lost Cause? A Historical Perspective on Conceptual Problems in Sentencing Research*, 76 Law & Contemp. Prob. 237 (2013).

[5]   Travis W. Franklin et al., *Extralegal Disparity in the Application of Intermediate Sanctions: An Analysis of U.S. District Courts*, 63 Crime & Delinq. 839, 840 (2017) (forthcoming) [hereinafter Franklin et al., *Intermediate Sanctions*].

[6]   Most studies rely upon single-level regression models. Jose Pina-Sánchez & Robin Linacre, *Refining the Measurement of Consistency in Sentencing: A Methodological Review*, 44 Int'l J. L. Crime & Just. 68, 78 tbl.1 (2016). For more information on the potential limitations on single-level models, see the methodological Appendix.

for more research on court-level factors in judicial decisionmaking.[7] In the federal system, individual defendants are nested (i.e., clustered) within groups at a higher level, being district courts. It is hypothesized that unique courtroom workgroups within district courts result in sentencing practices that differ across districts. Multilevel modeling, explained further herein, provides the ability to investigate how certain predictor factors are related to upward departures in individual cases while also testing whether the effects of those same factors differ among districts.

The Article proceeds as follows. Section II outlines the federal sentencing guidelines system. It then turns to upward departures specifically to contextualize the many reasons they represent extraordinary decision points worthy of scrutiny. Section III reviews contested issues concerning whether disparities are ever warranted and specifically addresses the challenge of regional disparities. Two theoretical views on disparities are relevant. The focal concerns perspective demonstrates that individual penalties tend to be based on perceptions of the defendant's culpability, the defendant's risk of recidivism, and the practical consequences of the potential punishment. In turn, the courtroom communities' perspective indicates that judges and practitioners in courtroom workgroups develop their own unique traditions and routines, which can explain some variations between courts in sentencing outcomes. Next, a literature review summarizes the results of prior empirical research on federal sentencing practices. The preexisting research was informative to building the statistical models presented herein.

Section IV sets forth an original empirical study of upward departure decisions. The data and variables are explained and the results from the multilevel models on upward departures are provided. In sum, the results demonstrate a statistically significant variance between district courts on upward departure outcomes. In a full model, a host of legal factors (e.g., final offense level, criminal history, offense type), extralegal characteristics (e.g., gender, race/ethnicity, citizenship), and case-processing variables (e.g., custody status) are predictive of upward departure outcomes in individual cases. Yet the influence of most of them varies across district courts, suggesting regional disparities in outcomes. The implications of the findings regarding factors correlated with individual outcomes and regional disparities are discussed in more detail. The results also substantively support the focal concerns and courtroom communities' perspectives. A methodological Appendix attached hereto further demonstrates the empirical benefits of a multilevel regression modeling approach and describes foundational decisions underlying the final results reported in the main text.

II. History and Current Guidelines Practices

This Article reports an original study using a sophisticated empirical modeling strategy to explore decisionmaking in criminal penalties. More specifically, the study is of discretionary upward departure outcomes in the federal sentencing system. A focus on criminal justice research specifically at the federal level is

---

[7]   Rob Tillyer & Richard Hartley, *The Use and Impact of Fast-Track Departures: Exploring Prosecutorial and Judicial Discretion in Federal Immigration Cases*, 62 Crime & Delinq. 1624, 1640 (2016).

meaningful for several key reasons. In contemporary times, federal authorities act as a role model in the administration of justice.

> [The federal government] provides resources, collects and develops best practices, and serves as the communicator and facilitator of these best practices throughout the country. . . . Because state, local, and tribal governments are limited by the need to devote resources to solving problems unique and endemic to their particular jurisdictions, the [f]ederal government plays [an] explicit role[] in advancing public policy to respond to gathering threats.[8]

Congress itself is often perceived as a leader in setting the criminal justice policy agenda for the country.[9] With respect to the federal government influencing sentencing decisions, the Justice Department at times has used funding programs to encourage states to adopt federally-based sentencing practices, such as determinate penalties and sentencing enhancements.[10] In addition, the federal sentencing guideline structure has been a model for the states who have adopted guideline systems.

Still, the federal guidelines are known for their extraordinary complexity[11] and are considered the most detailed[12] and constraining[13] ever developed in the country. The federal guidelines clearly were meant to restrain discretion in sentencing. The complex and detailed nature of the federal Guidelines mean that departures from them may provide particularly significant information about relevant predictors in this type of discretionary decisionmaking.[14] The potential to observe seeming disparities, even possibly implicit discrimination, is therefore informative to those interested in fairness, consistency, and transparency in decisions regarding punishments. Studies on federal sentencing also offer a benefit of representing judicial decisions across the country, thus perhaps making the results more generalizable than would research on a single state or subdivision of a state.

There is another significant way that the federal system has influence on the evolution of criminal justice responses in the country. In part due to what some critics perceive as overcriminalization in Congress' enactment of scores of new

---

[8] NAT'L CRIM. JUST. ASSOC, THE FEDERAL GOVERNMENT'S ROLE IN JUSTICE ADMINISTRATION 3 (2005), available at http://www.ncja.org/issues-and-legislation/role-federal-govt-administration-justice/role-federal-govt-administration.

[9] Jerold Israel, *Federal Influence in State Cases: Sentencing, Prosecution, and Procedure*, 543 ANNALS 130, 131 (1996).

[10] John F. Pfaff, *Federal Sentencing in the States: Some Thoughts on Federal Grants and State Imprisonment*, 66 HASTINGS L.J. 1567, 1571 (2015); Lisa L. Miller, *Looking for Post-Modernism in all the Wrong Places*, 41 BRIT. J. CRIMINOLOGY 168, 172 (2001).

[11] James C. Oleson et al., *The Sentencing Consequences of Federal Pretrial Supervision*, 63 CRIME & DELINQ. 313, 315 (2017).

[12] Paul J. Hofer et al., *The Effect of the Federal Sentencing Guidelines on Interjudge Sentencing Disparity*, 90 CRIMINOLOGY 239, 240 (1999) [hereinafter Hofer at al., *Sentencing Guidelines*].

[13] Ben Grunwald, *Questioning Blackmun's Thesis: Does Uniformity Sentencing Entail Unfairness*, 49 LAW & SOC'Y REV. 499, 500 (2015).

[14] Douglas A. Berman, *From Lawlessness to Too Much Law: Exploring the Risk of Disparity from Differences in Defense Counsel under Guidelines Sentencing*, 87 IOWA L. REV. 435, 445 (2002).

federal criminal laws over the last few decades,[15] the federal government now operates the single largest criminal justice system by inmate count in the United States.[16] Indeed, the federal prison system itself is among the top ten largest by country in the world.[17]

To situate the context of this study on upward departure decisions, a brief summary of the federal guidelines system is offered. Then the discussion outlines the case for why upward departures are noteworthy discretionary decisions that offer a valuable subject for research.

A. PRIMER ON FEDERAL GUIDELINES

At the turn of the twentieth century, the federal sentencing system represented an indeterminate structure that awarded federal district judges broad discretion to determine criminal penalties in individual cases.[18] By the 1970s, however, critics objected. Complainants alleged that the indeterminate structure led to unappealing results, such as too lenient sentences for certain offenses, disparities in sentences among similarly-situated offenders, and discrimination against minority defendants.[19] In its place, the country's politicians across the country embarked in the 1980s on a mission to enact more determinate policies.[20]

Congress was at the forefront of the country's reform movement in the latter part of the twentieth century by adopting legislation which mandated more regimented sentencing practices. The Sentencing Reform Act of 1984 created a presumptive sentencing system to be engineered under the auspices of a newly formed United States Sentencing Commission (the "Commission" or "Sentencing Commission").[21] A dramatic and holistic reform ordered the Commission develop a determinate system of sentencing guidelines ("Sentencing Guidelines" or "Guidelines") to systematize sentencing outcomes principally by restraining judicial discretion. "Proponents of this package hoped that it would end judge-to-judge and region-to-region disparities, promote candor in sentencing, and provide judges with relative values in sentences."[22]

---

[15] Rachel E. Barkow, *Federalism and Criminal Law: What the Feds can Learn from the States*, 109 MICH. L. REV. 519, 524-27 (2011).

[16] NATHAN JAMES, CONG. RES. SERV., R42937, THE FEDERAL PRISON POPULATION BUILDUP: OPTIONS FOR CONGRESS 1 (2016).

[17] U.S. SENT'G COMM'N, QUICK FACTS: FEDERAL OFFENDERS IN PRISON 1 (2015) (noting 210,567 inmates in federal prison as of February 2015, with 185,644 of them serving a federal sentence). The 185,644 figure just given represents the nine largest in the world following China, Russia, Brazil, India, Thailand, Mexico, Iran, and Turkey. *See* INT'L CENTRE FOR PRISON STUD., WORLD PRISON BRIEF http://www.prisonstudies.org/highest-to-lowest/prison-population-total?field_region_taxonomy_tid=All.

[18] Ilene H. Nagel, *Discretion: The New Federal Sentencing Guidelines*, 80 J. CRIM. L. & CRIMINOLOGY 883, 893 (1990).

[19] Kate Stith & Steve Y. Koh, *The Politics of Sentencing Reform: The Legislative History of the Federal Sentencing Guidelines*, 28 WAKE FOREST L. REV. 223, 227-28 (1993).

[20] Michael Tonry, *Sentencing in America, 1975–2025*, *in* 42 CRIME AND JUSTICE IN AMERICA, 1975–2025, at 141, 159–60 (Michael Tonry ed., 2013).

[21] Sentencing Reform Act of 1984, Pub. L. No. 98-473, §§ 211-300, 98 Stat. 1837, 1987-2040.

[22] Frank H. Easterbrook, *Introduction*, 26 AM. CRIM. L. REV. 1813, 1813 (1989).

An unforeseen and significant development recast how the Guidelines were to operate. Despite Congress' intent for a presumptive Guidelines system, the United States Supreme Court rendered the Guidelines advisory in nature. In the seminal case of *United States v. Booker* in 2005, the Court found that the system operated in an unconstitutional manner because judges, rather than juries, were the arbiters of facts that increased sentence length.[23] Bestowing advisory status was the Supreme Court's remedial fix to avoid overturning the entire Guidelines system.[24]

The *Booker* fix did not, however, return to the judiciary the wide discretion that existed pre-Guidelines. In a series of cases since then, the Supreme Court has reaffirmed that federal judges remain significantly circumscribed by the Commission's Guidelines and policies.[25]

At their heart, the Guidelines provide for a series of calculations in order to determine the defendant's offense severity level and criminal history score. With these two numbers in hand, the district judge consults a single Guidelines grid to obtain the recommended prison sentence.[26] The grid is not the end of the decisionmaking process though. Once the Guidelines-recommended penalty for the individual defendant is determined, the judge considers whether any departure provision contained in the Guidelines may apply.[27] Guidelines-based departures may be downward or upward, meaning either that they would justify a sentence below or above, respectively, from the recommendation. The Guidelines contain a number of provisions which the Commission staff acknowledges are circumstances that may not be adequately covered in the offense severity and criminal history provisions. Two of the downward departures expressly require the affirmative motion of the government to justify them.[28]

The Guidelines expressly provide for several types of upward departures, all of which are discretionary to the judge and do not require the prosecutor's request.[29] An example given for an approved upward departure (and one that is relevant to the results of the study provided herein) addresses the inadequacy of the computed criminal history category to properly reflect the defendant's deviant past.[30] Reasons specified for why the judge may find the official criminal history category inadequate include the existence of prior similar conduct not resulting in a criminal conviction or when a prior sentence was not officially computed in the criminal history calculation (e.g., the prior sentence was too dated and thus was excluded from the official calculation).[31]

---

[23] The Court ruled that such judicial factfinding violated the Sixth Amendment. 543 U.S. 220, 245 (2005).

[24] 543 U.S. at 249.

[25] Peugh v. United States, 186 L. Ed. 2d 84, 95 (2013).

[26] U.S. SENTENCING GUIDELINE MANUAL ch. 5 pt. A, sent'g tbl. (2015).

[27] *Id*. at § 1B1.1(b).

[28] These are substantial assistance to authorities in investigating another potential offender (§ 5K1.1) and fast-track departures as a docket-clearing option (§ 5K3.1).

[29] Technically, there are two types of upwardly varying sentences in the federal system. A "departure" is a term used in the Guidelines which refers to a sentence outside the recommended range from the sentencing grid but permitted by the Guidelines rules. United States v. Jeffers, 2015 U.S. Dist. LEXIS 132055, at *21-22 (N.D. Iowa 2015). A "variance" is a non-Guidelines sentence invoked to achieve statutory sentencing goals. *Id.* The difference between them is not of consequence here and the Article uses "upward departure" generally to signify both of them.

[30] U.S. SENTENCING GUIDELINE MANUAL § 4A1.3.

[31] *Id*. at § 4A1.3.

Per the statutory framework and Guidelines policy, a judge may also depart for reasons not included in the Guidelines if "there exists an aggravating or mitigating circumstance of a kind, or to a degree, not adequately taken into consideration by the Sentencing Commission in formulating the Guidelines."[32] Judges may reject the recommendation for other reasons, including, according to the Supreme Court in a case following *Booker*, based on a direct policy dispute with a relevant Guideline or Commission policy.[33] Nevertheless, the Guidelines preclude consideration of the defendant's race, sex, national origin, and socioeconomic status.[34]

In the end, a district judge in the individual case must determine a penalty that is reasonable and parsimonious, one that comprises "a sentence sufficient, but not greater than necessary."[35] The penultimate step, then, is for the judge to reflect upon whether a within-Guidelines or, alternatively, a non-Guidelines penalty is proper.[36] Then she pronounces the sentence.

The existence of greater discretion afforded by *Booker* have led empirical researchers to study how discretion is used and whether differences in sentencing outcomes across judges and districts may be a repercussion.[37] The study of potential disparities herein focuses on upward departure decisions for the reasons that are outlined next.

B. The Significance of Upward Departures

It is curious that there appear to be no other empirical studies comprehensively concentrating on upward departures in the federal system. Departures upward are extraordinary and consequential decisions for many reasons. First, an upward departure obviously is meant to increase the severity of the penalty. Prior studies in federal sentencing confirm such a result, and they demonstrate that the consequences are significant. Regression studies have found that the decision to upwardly depart multiplied the odds of a sentence involving incarceration by as much as 12 times compared to a sentence without an upward departure.[38] Regression results have also

---

[32] 18 U.S.C. § 3553(b) (2000); U.S. Sentencing Guideline Manual § 5K2.0 (2015).

[33] Kimbrough v. United States, 552 U.S. 85, 108-11 (2007).

[34] U.S. Sentencing Guideline Manual §4H1.10 (2015).

[35] 18 U.S.C. § 3553(a). The legislation specifies that district judges consider the following factors in determining a reasonable sentence in the individual case: (a) the recommended punishment range set by the sentencing guidelines and the Commission's policy statements; (b) the nature and circumstances of the offense; (c) the history and characteristics of the defendant; (d) the need for the sentence imposed considering the seriousness of the offense, retribution, deterrence, protecting the public, and the offender's rehabilitative needs; and (e) the need to avoid unwarranted sentencing disparities among similarly-situated offenders. *Id*.

[36] Gall v. United States, 552 U.S. 38, 50 (2007).

[37] William Rhodes et al., Federal Sentencing Disparity: 2005-2012, at *6 (2015).

[38] Brian D. Johnson & Sara Betsinger, *Punishing the "Model Minority": Asian-American Criminal Sentencing Outcomes in Federal District Courts*, 47 Criminology 1045, 1067 tbl. 3 (2009). *See also* Travis W. Franklin, *Sentencing Outcomes in U.S. District Courts: Can Offenders' Educational Attainment Guard against Prevalent Criminal Stereotypes*, 36 Crime & Delinq. 137, 151 tbl. 2 (2017) [hereinafter Franklin, *Educational Attainment*] (finding upward departures increased the odds of incarceration by 11 times and increased sentence length by 83%); Travis W. Franklin, *Sentencing Native Americans*

indicated that an upward departure as much as doubles the length of the resulting prison sentence.[39]

Second, to the extent that upward departures naturally leads to a greater number of defendants being incarcerated and for longer periods, these decisions worsen the federal system's prison overpopulation problem. Since 1980, the federal prison population has grown 750%.[40] As a result, the federal prison system is challenged by the resulting increases in costs of imprisonment and is dangerously overcrowded.[41] An Urban Institute report has tagged longer sentences as contributing to over half of the growth in the federal prison system.[42] Upward departure outcomes—whether considered legitimate or not—exacerbate these tensions.

Third, upward departures uniquely signal that judges may be finding gaps in Guidelines policies and calculations, despite the Commission's now decades of experience with studying sentencing practices and making relevant policy adjustments as needed. When a judge determines whether to depart upward from the Guidelines recommendation, it likely represents a compromise between uniformity and proportionality. Whereas downward departures are often for reasons other than proportionality concerns (for example, the repeated use of fast-track departures and substantial assistance departures are mainly for efficient case-processing purposes), upward departures are more attuned to calibrating the penalty to the defendant's culpability and harm. Upward departures are even more surprising as many judges, practitioners, and researchers already assess the Guidelines as producing excessively harsh sentence recommendations as a general rule.[43] Thus, upward departures appear to be exceptions to the rule about the sufficiency (or tendency toward excessiveness) of Guidelines-based proportionality judgments.

Fourth, because upward departures are relatively rare, it is therefore even more symbolic when one is issued in an individual case.[44] An upward departure constitutes individualized sentencing since it is an ad hoc, discretionary decision. The rare

---

[39] *in US Federal Courts: An Examination of Disparity*, 30 Just. Q. 310, 326 tbl. 2 (2013) (finding upward departures increased the odds of incarceration by a factor of seven).

[39] Tillyer & Hartley, *supra* note 7, at 1635 tbl. 2 (obtained by anti-logging the coefficient of .71); Jeffery Ulmer & Michael T. Light, *Beyond Disparity: Changes in Federal Sentencing After* Booker *and* Gall*?*, 23 Fed. Sent'g Rep. 333, 336 tbl. 2 (2011); Ben Feldmeyer & Jeffery T. Ulmer, *Racial/Ethnic Threat and Federal Sentencing*, 48 J. Res. Crime & Delinq. 238, 252 tbl. 3 (2011); Celesta A. Albonetti & Robert D. Baller, *Sentencing in Federal Drug Trafficking/Manufacturing Cases: A Multilevel Analysis of Extra-Legal Defendant Characteristics, Guidelines Departures, and Continuity of Culture*, 14 J. Gender Race & Just. 41, 68 tbl. 3 (2010) (studying drug trafficking cases).

[40] Samuel A. Taxy, Drivers of Growth in the Federal Prison Population 1 (2015), available at http://www.urban.org/research/publication/drivers-growth-federal-prison-population.

[41] *See generally* Nathan James, Cong. Res. Serv., R42937, The Federal Prison Population Buildup: Options for Congress (2016).

[42] Kamala Mallik-Kane et al., Examining Growth in the Federal Prison Population, 1998 to 2010, at *10 (2012), available at http://www.urban.org/research/publication/examining-growth-federal-prison-population-1998-2010.

[43] Byungbae Kim et al., *The Impact of* United States v. Booker *and* Gall/Kimbrough v. United States *on Sentence Severity: Assessing Social Context and Judicial Discretion*, 62 Crime & Delinq. 1072, 1075 (2016); Cassia Spohn, *Twentieth-Century Sentencing Reform Movement: Looking Backward, Moving Forward*, 13 Criminology & Pub. Pol'y 535, 538 (2014).

[44] Upward departures occur in three percent of cases. Data obtained from the Commission's annual sourcebooks.

upward departure may, then, be acutely felt as unforeseeable and unfair, perhaps even arbitrary. These perceptions challenge the integrity of the system. Notably, a judge issuing a sentence that constitutes an upward departure does not do so by mistake or in ignorance. The Commission requires district courts to complete a Statement of Reasons form for each sentence which includes several fields where an upward departure box must be checked (when applicable) and further justified.[45]

An upward departure is also a particularly risky choice. In part because of its rarity and in part because of the substantive due process rights afforded criminal defendants, an upward departure practically invites the defendant to appeal. On review, the upward departure decision may well be overturned, particularly if the appellate court finds that the district judge did not provide sufficient reasons for the higher sentence.[46]

Fifth, upward departures are surprising, too, as they violate the premise underlying the cognitive bias of anchoring.[47] Anchoring effects refer to a person's tendency when making numbers-based judgments to rely on numeric reference points.[48] Anchoring is an example of a psychological heuristic in providing a shortcut to more efficient decisionmaking by tuning the person's thought process toward the given anchor number.[49] The Guidelines are generally considered to be substantive anchors for sentencing decisions.[50] An upward departure, then, requires the particular judge to reject the anchor and thereby lose the value of the cognitive shortcut. A discretionary decision to depart imposes a further resource cost upon the judge issuing it because of the burden to justify it in writing in the Statement of Reasons and in a way that distinguishes the case from the heartland already covered by the Guidelines.[51]

Sixth, it is widely recognized that departure decisions as a general rule (upward and downward) are significant, if not primary, sources of perceived disparities in

---

[45] *See generally* Jelani Jefferson Exum & Paul J. Hofer, *The Evolution of the Statement of Reasons Form*, 28 Fed. Sent'g Rep. 169 (2016).

[46] *See e.g.*, United States v. Howard, 773 F.3d 519 (4th Cir. 2014) (vacating upward departure as district court's judgment about defendant's criminal past insufficient to support it); United States v. Espinoza, 550 Fed. Appx. 690 (11th Cir. 2013) (vacating upward departure as district court did not adequately justify it); United States v. Conroy, 567 F.3d 174 (5th Cir. 2009) (vacating upward departure as district judge erred in analyzing whether the defendant's conduct met the Guidelines-based departure provision); United States v. Dillon, 355 Fed. Appx. 732 (4th Cir. 2009) (remanding sentence a second time as sentencing judge did not adequately explain its justification); United States v. Ofray-Campos, 534 F.3d 1 (1st Cir. 2008) (reversing upward variance where reasons given not compelling enough for an extraordinary variance).

[47] Silvio Aldrovandia et al., *Sentencing, Severity, and Social Norms: A Rank-Based Model of Contextual Influence on Judgments of Crimes and Punishments*, 144 Acta Pschyologica 538, 546 (2013).

[48] Jeffrey J. Rachlinski et al., *Can Judges Make Reliable Numerical Judgments: Distorted Damages and Skewed Sentences*, 90 Ind. L.J. 695, 695 (2015).

[49] Bettina von Helversen & Jörg Rieskamp, *Predicting Sentencing for Low-Level Crimes: Comparing Models of Human Judgment*, 15 J. Experimental Psychol. 375, 379 (2009).

[50] *See generally* Melissa Hamilton, *Extreme Prison Sentences: Legal and Normative Consequences*, 38 Cardozo L. Rev. 59 (2016) [hereinafter Hamilton, *Extreme Sentences*] (reviewing literature on anchoring effects, providing an empirical study on anchoring effects of Guidelines recommendations on sentencing outcomes, and concluding anchoring exists in federal sentencing practices).

[51] *See* Andrew W. Nutting, *The* Booker *Decision and Discrimination in Federal Criminal Sentences*, 51 Econ. Inquiry 637, 641 (2013).

sentencing.[52] If judges depart from Guidelines recommendations too often or for inappropriate reasons, they may be thwarting the main purpose of the implementation of the Guidelines system of reducing unwarranted disparities.[53] Upward departures, unlike some downward departures, do not require a prosecutorial motion, and thereby provide a mechanism for which judicial discretion unequivocally impacts sentencing severity. Plus, when such discretion is based on extralegal (i.e., not legally or formally permissible) reasons, the resulting judgments may even implicate implicit race, gender, or class discrimination. Importantly, researchers have previously tied extralegal factors to decisions that deviate from the Guidelines.[54]

This suggested relationship between upward departures and discretion is highlighted by the likely impact of the *Booker* decision (granting judges greater discretionary ability) on the rate of upward departures. The year after *Booker*, the rate of upward departures doubled compared to the annual rate of upward departures in the decade preceding the decision.[55] The rate of upward departures is now (i.e., fiscal years 2014-2015) at three times the pre-*Booker* rate.[56] Since the *Booker* decision (through the end of fiscal year 2015), federal judges have upwardly departed from Guidelines' recommendations in over 15,000 cases.[57] As another empirical verification of the role of discretion (possibly even discrimination), a substantial majority of these upward departures after *Booker*, as reported by judges themselves in the Statement of Reasons, are based on grounds other than the upward departure policies explicitly permitted by the Guidelines.[58]

Thus far, it has been argued that upward departures in federal sentencing are worthy of further analysis. The study was also led by relevant normative and theoretical foundations and informed by the results of previous studies.

III. Normative, Theoretical, and Research Consideration

The issue of disparities in sentencing practices is not a simple concept and not all agree on either whether it is necessarily a bad result. Challenges presented by potential disparities in penalties are discussed next. Then the Section reviews two major

---

[52] Jawjeong Wu & Cassia Spohn, *Interdistrict Disparity in Sentencing in Three U.S. District Courts*, 56 Crime & Delinq. 290, 296-97 (2010); Brian D. Johnson, Jeffery T. Ulmer, & John H. Kramer, *The Social Context of Guidelines Circumvention: The Case of Federal District Courts*, 46 Criminology 737, 740 (2008) [hereinafter Johnson et al., *Social Context*]; Hofer at al., *Sentencing Guidelines*, *supra* note 12, at 240.

[53] Michael S. Gelacak et al., *Departures Under the Federal Sentencing Guidelines: An Empirical and Jurisprudential Analysis*, 81 Minn. L. Rev. 299, 303 (1996).

[54] Jeffery T. Ulmer et al., *Racial Disparity in the Wake of the* Booker/Fanfan *Decision: An Alternative Analysis to the USSC's 2010 Report*, 10 Criminology & Pub. Pol'y 1077, 1080 (2011); Johnson et al., *Social Context*, *supra* note 52, at 740.

[55] Data analyses done by author using the Commission's data files from fiscal years 1999-2015 and the Commission's annual Sourcebooks for fiscal years 1989-2015.

[56] U.S. Sentencing Commission, 2015 Sourcebook tbl. N; 2008 Sourcebook tbl. N; 2006 Sourcebook tbl. N.

[57] Results from the author's frequency distribution analysis run of the Commission's datasets.

[58] The conclusion is derived from the Commission's annual Sourcebooks from fiscal years 2008-2015.

theoretical viewpoints relevant to the research herein, which are referred to as the focal concerns perspective and the courtroom workgroup perspective. Following that is a concise empirical literature review of relevant studies of federal sentencing practices.

A. DISPARITY ISSUES

The Sentencing Commission clearly values national uniformity in case-processing and outcomes.[59] While the tenets of federalism philosophically permit criminal laws to vary by state, federal criminal law is expected to provide a single set of policies regarding the official reaction to offenders who commit crimes that are of national interest.[60] Guidelines are expressly meant to provide a normative function.[61] Indeed, the federal Guidelines have over their thirty year existence become embedded in the legal, political, and organizational cultures of federal court communities.[62]

The Commission is not the only institution that works to normalize federal sentencing practices across judicial districts. The U.S. Department of Justice and the Federal Judicial Center are also centralized authorities providing educational opportunities to socialize judges into the federal government's sentencing policies.[63] Offering frequent training in the form of written primers, face-to-face instructional classes, and web-based videos[64] are necessary because of the complexity of the Guidelines. The 2015 Guidelines Manual is just shy of 600 pages,[65] with hundreds, if not thousands, of rules, depending on how one parses the rule counting scheme. The unavoidable purpose for such complexity is to try to leave as little uncovered as possible and thus to correct for potential lapses. Consistent with such intent, the Commission asserts that the primary goal of the sentencing Guidelines was to "eliminate" (i.e., implying not just reduce) unwarranted sentencing disparities.[66]

Though not all stakeholders would concur, it is not always clear what disparity means and whether it is necessarily a bad thing. According to Black's Law Dictionary, disparity means "inequality" and "a difference in quantity or quality between two or more things."[67] The first meaning (inequality) tends to have a negative connotation, at least in criminal justice circumstances. The second (oriented around differences)

---

59 U.S. SENTENCING GUIDELINE MANUAL ch. I, Pt. A, at 1.3 (2015) ("Congress sought reasonable uniformity in sentencing by narrowing the wide disparity in sentences imposed for similar criminal offenses committed by similar offenders."); U.S. SENTENCING COMMISSION, 2014 ANNUAL REPORT A-3 (2014).

60 Stephanos Bibas, *Regulating Local Variations in Federal Sentencing*, 58 STAN. L. REV. 137, 137 (2005).

61 RHODES ET AL., *supra* note 37, at 23 n. 19.

62 Ulmer & Light, *supra* note 39, at 340.

63 Jeffery T. Ulmer, *The Localized Uses of Federal Sentencing Guidelines in Four U.S. District Courts: Evidence of Processual Order*, 28 SYMBOLIC INTERACTION 255, 256-57 (2005) [hereinafter Ulmer, *Localized Uses*].

64 For a glimpse into the various instructional offerings, see the Commission's training website: http://www.ussc.gov/topic/training.

65 *See generally* U.S. SENTENCING GUIDELINE MANUAL (2015).

66 U.S. SENT'G COMM'N, FIFTEEN YEARS OF GUIDELINES SENTENCING: AN ASSESSMENT OF HOW WELL THE FEDERAL CRIMINAL JUSTICE SYSTEM IS ACHIEVING THE GOALS OF SENTENCING REFORM 79 (2004) [hereinafter FIFTEEN YEARS].

67 BLACK'S LAW DICTIONARY (10th ed. 2014).

does not necessarily carry an adverse inference. Such competing alternatives to the implication of using the term disparity similarly complicates the discussion in criminal justice circles.

When observers discuss disparity in sentencing outcomes, it is often based on identifying like individuals who commit like offenses.[68] Disparity in this sense might be viewed as the flipside of uniformity in which the posited individuals received similar punishments. An obvious critique of these philosophical notions is that there is no objective criteria for determining what exactly constitutes *like* individuals or *like* offenses. With the complexity of human nature and conduct, no individual or deed can truly be identical.

In any event, the Guidelines—despite *Booker*—remain the lodestone of federal sentencing practices.[69] Still, many sources are again concerned with perceived disparities in actual sentencing decisions.[70] What do they tend to consider is wrong with disparities in punishment? Rationales are that differences in punishment for like offenses erodes the public confidence in an expectedly legal, objective, and rational system,[71] and that they bring gratuitous uncertainty and unfairness[72] for defendants, victims, the government, and the public.

The posited problems with disparities are particularly acute when judges base sentences on extralegal factors that the Guidelines were intended to more proactively forbid.[73] Some argue that empirical evidence of differential sentencing practices based on demographic factors is obviously indicative of illegal discrimination.[74] Their issue is not just with overtly discriminatory practices. The *Booker* decision increased ambiguity in the exact reasons for district court decisions and thereby multiplied the potential for implicit discrimination, meaning unconscious and unintentional discrimination in individual cases.[75] Thus, implicit discrimination might arguably be present when studies show that females and whites, for instance, routinely receive lesser punishments than males and blacks, respectively, after controlling for relevant legal factors.[76] Variations in sentencing practices may be signs not only of inequality and injustice, they also undermine the deterrence value of predictable and firm sentencing policies.[77]

Nonetheless, it is still reasonable to acknowledge that not all variances from Guidelines recommendations constitute disparities, particularly in the negative sense of the term. Prior statisticians reviewing federal sentencing data rightly observe that a non-Guidelines-compliant sentence is not necessarily illegal considering the discretion that judges now lawfully maintain to deviate per *Booker*.[78] Further, as

---

[68] RHODES ET AL., *supra* note 37, at 7.

[69] U.S. SENT'G COMM'N, FEDERAL SENTENCING: THE BASICS 3 (2015).

[70] U.S. SENT'G COMM'N, 2012 REPORT TO THE CONGRESS: CONTINUING IMPACT OF UNITED STATES V. BOOKER ON FEDERAL SENTENCING F-9 (2012) (citing sources).

[71] Mandeep K. Dhami et al., *Quasirational Models of Sentencing*, 4 J. APPLIED RES. MEMORY & COGNITION 239, 242 (2015).

[72] FIFTEEN YEARS, *supra* note 66, at 38.

[73] J.C. Oleson, *Blowing out the Candles: A Few Thoughts on the Twenty-Fifth Anniversary of the Sentencing Reform Act of 1984*, 45 RICH. L. REV. 693, 755 (2011).

[74] Pina-Sánchez & Linacre, *supra* note 6, at 72; Hofer at al., *Sentencing Guidelines*, *supra* note 12, at 242.

[75] Nutting, *supra* note 51, at 638-39.

[76] *Id*. at 645.

[77] Bibas, *supra* note 60, at 137.

[78] RHODES ET AL., *supra* note 37, at 18.

an appellate judge reasonably stated, "while a strictly code-based method of legal problem-solving might work to achieve predictability and some sort of uniformity, it does not always work to achieve justice."[79] The inability or unwillingness of a judge to depart from the Guidelines may inequitably mean there is an inordinate amount of rigidity in sentencing requirements.[80] Hence, a reciprocal danger of unwarranted disparity to notions of justice is unwarranted uniformity.

There may well be something extraordinary in a particular case where a judge's discretionary ability could work to better serve justice for all parties.[81] Some commentators thus point out the desirability of individualizing penalties.[82] Likely, balancing is the key. There is some value in providing judges some discretionary ability in determining penalties to account for exceptional circumstances, even if there is also value in channeling or controlling that discretion to avoid abuses.[83]

In the end, this paper does not take the concrete position that even sophisticated statistical analyses of sentencing outcomes can prove that every upward departure represents disparity, at least to the extent the term holds a negative connotation, much less a discriminatory decision. Nor does the paper assign condemnatory blame to district judges for differences in sentencing for seemingly comparable offenses or offenders. As with any study of human behavior, no dataset can possibly account for all aspects of criminal conduct or of decisionmaking. Thus, different judges may sentence seemingly similar offenders to incomparable punishments for legitimate reasons that are simply not captured in the data.

Further, the source of any unwarranted disparity may arise from other actors anyway, such as based on the (legitimate or illegitimate) practices and decisions of other actors in the criminal justice process chain.[84] Research has shown that prosecutors can finesse facts in their case filings and to manipulate the offense(s) charged and/or the specific offense characteristics on which the Guidelines computation is based.[85] Contributions to differences in sentencing outcomes may also derive from inconsistent policies in policing or in the preparation of presentence reports by probation officers.[86] Disparities in outcomes for otherwise seemingly similar offenders may likewise depend upon the diverse competencies of defense counsel with respect to their grasp of the complex Guidelines system.[87]

---

[79] Rosemary Barkett, *Judicial Discretion and Judicious Deliberation*, 59 Fla. L. Rev. 905, 918 (2007).

[80] Michael S. Gelacak et al., *Departures Under the Federal Sentencing Guidelines: An Empirical and Jurisprudential Analysis*, 81 Minn. L. Rev. 299, 303 (1996).

[81] Paul J. Hofer, United States v. Booker *as a Natural Experiment: Using Empirical Research to Inform the Federal Sentencing Policy Debate*, 6 Criminology & Pub. Pol'y 433, 438-39 (2007).

[82] Amy Baron-Evans & Kate Stith, Booker *Rules*, 160 U. Penn. L. Rev. 1631, 1648 (2012); W.H. Townsend, *The Punishment of Crime*, 10 J. Am. Inst. Crime & Criminology 533, 535 (1920) ("Individualization is the process of adjusting a penalty to the character of a criminal. The criterion of judgment is threefold, including the crime, social conditions, and the criminal.").

[83] Stuart S. Nagel, *Discretion in the Criminal Justice System: Analyzing, Channeling, Reducing and Controlling It*, 31 Emory L.J. 603, 609 (1982).

[84] Besiki L. Kutateladze et al., *Cumulative Disadvantage: Examining Racial and Ethnic Disparity in Prosecution and Sentencing*, 52 Criminology 514, 517 (2014).

[85] Rhodes et al., *supra* note 37, at 7.

[86] Fifteen Years, *supra* note 66, at 84.

[87] Douglas A. Berman, *From Lawlessness to Too Much Law*, 87 Iowa L. Rev. 435, 445 (2002).

Despite the choice not to assume all differences in outcomes establish unwarranted disparities, the observation that "some *patterns* in those differences are suggestive of disparity"[88] in its more negative sense appears reasonable. What the study herein can do is to parse the patterns of differences in the outcomes of upward departures (versus not) that might imply these disparities.

B. REGIONAL DIFFERENCES

Another disparity matter needs to be addressed considering the study contained herein will focus on it: regional variations in sentencing outcomes. The issue here is where sentencing outcomes may be uniformly meted out within a region but vary from those in other regions. Regional disparities are viewed by some observers in unfavorable terms. The Sentencing Commission officially asserts that the federal Guidelines were meant to control local variations in sentencing practices, such that consistent practices were intended to be enforced nationwide when prosecuting federal crimes.[89] A few commentators agree that any regional disparities for local concerns are necessarily extralegal in nature and thus indefensible and that, because they are *extra*legal, their sheer existence nullifies a major purpose of the Guidelines.[90]

Before reviewing potential sources of regional differences in federal sentencing outcomes, two limitations in the study's design should be noted here. Federal district courts are comprised of more than one district judge.[91] As each sentencing decision is the product of a single judge, a preferable method would be to study interjudge outcomes. However, the Sentencing Commission deletes judge identifiers from its datasets such that it was not possible to distinguish between individual judges within districts. Nonetheless, as judges within the same district may share more correlated characteristics than with judges from other district courts and as districts are regionally oriented, investigating district level disparities remains important. The datasets likewise do not include identifiers for probation officers or the recommended sentences listed in their authored presentencing reports.

There exist several potential sources of local variations in federal sentencing outcomes. One is that even though federal criminal law provides a single body of statutes covering the country equally,[92] federal district courts still are situated in fixed, single locales. Districts, thus, represent regions. Federal law may have nationwide coverage but the commission of federal crimes is not equally spread out across the country. Nor will victims of federal crimes in different areas necessarily experience their losses the same. A particular region might become a hotspot for gun violence related to drug trafficking while the citizens of another feel more acutely

---

[88] RHODES ET AL., *supra* note 37, at 18 (emphasis in original).

[89] FIFTEEN YEARS, *supra* note 66, at 90.

[90] Paula M. Kautt, *Location, Location, Location: Interdistrict and Intercircuit Variation in Sentencing Outcomes for Federal Drug-Trafficking Offenses*, 19 JUST. Q. 633, 635 (2002); Hofer at al., *Sentencing Guidelines*, *supra* note 12, at 243.

[91] *See generally* U.S. COURTS, CHRONOLOGICAL HISTORY OF AUTHORIZED JUDGESHIPS IN U.S. DISTRICT COURTS (2015), available at http://www.uscourts.gov/judges-judgeships/autho-rized-judgeships.

[92] This reference excludes criminal laws solely focused on the District of Columbia, native American lands, and federal property.

the negative impact of financial fraud. There may be some value in allowing judges to equitably adapt national policy to more localized concerns such as these, albeit in moderation.[93] Local variations may be proper, for instance, to swiftly and harshly respond to the area's particular crime problem, such as a district court increasing the severity of punishment for weapons offenses as a deterrent device to try to counter a rise in local gun violence. Such a strategy would obviously differentiate that court's sentencing statistics for firearm offenses.

Another possibility for regional variations is if there is local hostility to a national policy concerning a particular crime or the Commission's assessment of the severity of a crime. Observers may debate the propriety of a district judge's ability to void a centralized policy. Such a rationale may be viewed reasonably in culturally sensitive terms to accommodate local priorities or, instead, as an inappropriate usurpation of the lawful powers of federal policymakers to make national policy decisions.[94]

Other regional variations amongst federal courts in sentencing may be more or less benign, simply reflecting localized socialization in what are called courtroom workgroups. A cultural consensus unique to a courtroom workgroup may mean consistency in sentencing within that workgroup, but whose outcomes are uncorrelated (i.e., disparate) with outcomes generated by other courtrooms. This idea will be discussed further in the next Section that addresses two main theoretical foundations for between-court differences in criminal justice outcomes: the focal concerns perspective and the consequences of culturalized practices through the development of courtroom communities. For now, it is simply noted that the Sentencing Commission avers that regional variation in sentencing outcomes due to differing political climates or court cultures constitutes unwarranted disparity.[95]

C. Theoretical Foundations of Sentencing Decisions

The focal concerns perspective is now a popular theoretical framework for understanding sentencing outcomes.[96] The theory posits that decisions about penalties center on the authority's situational assessment concerning three focal concerns: (1) the defendant's culpability, (2) the defendant's future dangerousness, and (3) the practical consequences of the decision to the defendant and the community.[97]

The Guidelines certainly address the focal concerns in their formalized rules regarding assessments of blameworthiness (e.g., offense level representing severity, offense type), future dangerousness (e.g., criminal history, acceptance of responsibility), and consequences of the penalty (e.g., substantial assistance reductions to conserve prosecutorial resources, fast-track departures to permit more efficient case processing). Yet, considering human nature cannot always be entirely automated and the potential for highly-educated and experienced federal judges to believe in their own qualities of judgment, the Guidelines likely do not entirely constrain discretion in considering the focal concerns.

[93] Bibas, *supra* note 60, at 138.
[94] *Id*. at 140.
[95] Fifteen Years, *supra* note 66, at 80.
[96] Kutateladze et al., *supra* note 84, at 518.
[97] Jeffery T. Ulmer, James Eisenstein, & Brian D. Johnson, *Trial Penalties in Federal Sentencing: Extra-Guidelines Factors and District Variation*, 27 Just. Q. 560, 565-66 (2010).

Upward departures may rely more heavily on discretionary thought in that judges issuing them may be considering ideals or values not explicitly contained in the Guidelines rules. In addition, departure decisions beyond those expressed in the Guidelines presumably represent gaps in their set of rules. Thus, it is expected from the focal concerns perspective that there will be disparities in upward departure outcomes because of differences in judges' situational assessment of the focal concerns in individual cases, the extent of their agreement with the Guidelines-driven proportionality judgment, and their relative concern about the practical consequences of the sentence.

The second theoretical perspective popular in sentencing research regards community courtroom cultures. "Court communities are distinct, localized social worlds with their own relationship networks, organizational culture, political arrangements, and the like. These localized social worlds, with their organizational cultures and political realities, shape formal and informal case processing and sentencing norms."[98] Prior research consistently indicates that the type of sentence issued (e.g., probation versus imprisonment), the length of supervision, and the reasons for the particular penalty depend in part on the jurisdiction in which the defendant is sentenced because of localized differences in cultural, political, and social contexts.[99] Contextual variations in these court communities may result from the "participants' shared workplace and interdependent working relations between key sponsoring agencies (prosecutor's office, bench, defense bar)."[100] The courtroom community workgroup likely shares common experiences, and works together to develop normative practices to reduce uncertainty and serve a communal goal of efficient case processing.[101]

Empirical researchers tend to assume there exists little interdistrict variation in the federal system, specifically, because of the uniform set of laws and policies provided by federal statutes and the sentencing Guidelines.[102] As a result, interdistrict variations in penalties at the federal level are understudied simply because of the presumption of little variance.[103] This assumption is likely invalid as other observers contend that federal courts do not necessarily act with uniformity.

> We view the federal district court system not as a singular national legal structure with hierarchically arranged and geographically dispersed subunits, but rather as a semi-autonomous set of systems governed by the same formal rules, states, and procedural policies, while also embedded in localized legal cultures that are themselves shaped by regionally specific historical contingencies and norms.[104]

---

[98] Jeffery T. Ulmer & Mindy S. Bradley, *Variations in Trial Penalties among Serious Violent Offense*s, 44 Criminology 631, 641 (2006).

[99] Robert R. Weidner et al., *The Impact of Contextual Factors on the Decision to Imprison in Large Urban Jurisdictions: A Multilevel Analysis*, 51 Crime & Delinq. 400, 418 (2005).

[100] Ulmer & Bradley, *supra* note 98, at 641.

[101] Brian D. Johnson & Stephanie M. Dipietro, *The Power of Diversion: Intermediate Sanctions and Sentencing Disparity Under Presumptive Guidelines*, 50 Criminology 811, 819 (2012); Patricia D. Breen, *The Trial Penalty and Jury Sentencing: A Study of Air Force Courts-Martial*, 8 J. Empirical Legal Stud. 206, 213 (2011).

[102] Wu & Spohn, *supra* note 52, at 291-92.

[103] Johnson et al., *Social Context*, *supra* note 52, at 740.

[104] Mona Lynch & Marisa Omori, *Legal Change and Sentencing Norms in the Wake of* Booker: *The Impact of Time and Place on Drug Trafficking Cases in Federal Court,* 48 Law & Soc'y Rev. 411, 412 (2014).

Even though federal district courts operate at the national level, the practitioners within them are often plucked from their own locales. Idiosyncratic local practices within district court communities can impact federal sentencing as judges and prosecutors are often chosen from within the state in which the district court resides; plus, defense counsel and probation staff tend to have previously resided in or near the districts in which they become employed.[105] The Sentencing Commission does not discount the possibility of localized cultures. The agency has called for more lively research on geographic variations in sentencing practices and outcomes.[106] This Article responds to this call, too. The study herein was informed, as well, by previous empirical studies as to the most likely factors to consider in explaining federal sentencing outcomes.

D. LITERATURE REVIEW OF FEDERAL SENTENCING PRACTICES

Criminologists have aptly recognized that "offenders are sanctioned partially for what they have done (offense characteristics, criminal history), for who they are (race/ethnicity, age, gender) and also for what they may fail to do during the punishment process (plead guilty or express remorse)."[107] Researchers commonly refer to these considerations as representing legal factors, extralegal factors, and case-processing factors. They are consistent with the focal concerns perspective regarding culpability, risk, and external consequences to the punishment. Prior research on federal sentencing outcomes has tended to corroborate these sentiments. The United States Sentencing Commission undertakes a laudable effort to make available its rich datasets to researchers. This sub-section will summarize results from prior empirical studies on federal penalties which have utilized Commission datasets. The results provided necessary information on which variables this study tested as likely to be significant predictors of sentencing outcomes.

1. Significant Predictors of Sentencing Outcomes

As for legal factors, prior research has confirmed that primary predictors of federal sentencing outcomes are offense seriousness, criminal history,[108] and crime type.[109] As might be expected, multiple counts of conviction[110] and the application of a

---

[105] Michael Tonry, *Federal Sentencing "Reform" Since 1984: The Awful as Enemy of the Good*, 44 CRIME & JUST. 99, 124 (2015).

[106] FIFTEEN YEARS, *supra* note 66, at 112.

[107] Ronald S. Everett & Roger A. Wotkiewicz, *Difference, Disparity, and Race/Ethnic Bias in Federal Sentencing*, 18 J. QUANTITATIVE CRIMINOLOGY 189, 208 (2002).

[108] *E.g.*, Oleson et al., *supra* note 11, at 323 tbl. 2; Rob Tillyer et al., *Differential Treatment of Female Defendants: Does Criminal History Moderate the Effect of Gender on Sentence Length in Federal Narcotics Cases*, 42 CRIM. JUST. & BEHAV. 703, 705 (2015) [hereinafter Tillyer et al., *Gender*] (citing studies); Jeffery T. Ulmer et al., *Trial Penalties in Federal Sentencing: Extra-Guidelines Factors and District Variation*, 27 JUST. Q. 560, 576 tbl. 2 (2010) [hereinafter Ulmer et al., *Trial Penalties*].

[109] *E.g.*, Franklin, *Educational Attainment*, *supra* note 38, at 151 tbl. 2; Kim et al., *supra* note 43, at tbl. 2; Johnson et al., *Social Context supra* note 52, at 761 tbl. 5.

[110] *E.g.*, Jill K. Doerner & Stephen Demuth, *Gender and Sentencing in the Federal Courts: Are Women Treated More Leniently*, 25 CRIM. JUST. POL'Y REV. 242, 255 tbl. 2 (2014);

mandatory minimum sentence are associated with longer federal sentences.[111] In addition, official credit in the form of a reduction in offense levels for the defendant's acceptance of responsibility reduces sentence length in statistical models.[112]

Much research has found that demographic characteristics, which are generally considered to be extralegal factors for punishment purposes, are still correlated with sentence length. As for race and ethnicity, multiple studies of federal sentencing show that whites receive sentences of shorter length than blacks[113] and Hispanics even when controlling for various factors.[114] Several other projects find that the differences demonstrate unassailable racial disparities in federal sentencing.[115] A commonly applied theoretical explanation for assigning more severe penalties to racial and ethnic minorities relates to the minority threat thesis in which stereotypes of minorities being more likely to recidivate may enter into the focal concern of future dangerousness.[116]

Studies of sentencing rather consistently indicate that males are sentenced to longer periods of incarceration.[117] An explanation for the gender effect regards the

---

Kautt, *supra* note 90, at 655 tbl. 4 (studying drug offenses).

[111] *E.g.*, Kim et al., *supra* note 43, at 1084 tbl. 2; Lynch & Omori, *supra* note 104, at 432 (studying drug trafficking cases); Kautt, *supra* note 90, at 655 tbl. 4 (studying drug offenses). *See also* Melissa Hamilton, *Some Facts About Life: The Law, Theory, and Practice of Life Sentences*, 20 LEWIS & CLARK L. REV. 803, 848 tbl. 4 (2016) [hereinafter Hamilton, *Life Sentences*] (finding application of mandatory minimum predicted a sentence of 470 months or more).

[112] Feldmeyer & Ulmer, *supra* note 39, at 252 tbl. 3 (finding any acceptance of responsibility credit reduced sentence length 15%); Ulmer et al., *Trial Penalties, supra* note 108, at 576 tbl. 2 (finding each point reduction given for acceptance of responsibility reduced the sentence length 1% in model 2); Ulmer, *Localized Uses*, *supra* note 63, at 271 (finding acceptance of responsibility on average reduces sentences by a year in three of the districts studied).

[113] *E.g.*, Oleson et al., *supra* note 11, at 323 tbl. 2; Doerner & Demuth, *supra* note 110, at 255 tbl. 2; Feldmeyer & Ulmer, *supra* note 39, at 252 tbl. 3; Amy Farrell et al., *Intersections of Gender and Race in Federal Sentencing: Examining Court Contexts and the Effects of Representative Court Authorities*, 14 J. GENDER RACE & JUST. 85, 115 tbl. 3 (2010).

[114] Kim et al., *supra* note 43, at 1084 tbl. 2. *See also* Johnson et al., *Social Context*, *supra* note 52, at 761 tbl. 5 (finding whites more likely to receive downward departures than blacks and Hispanics).

[115] *E.g.,* Lynch & Omori, *supra* note 104, at 432 (studying drug offenders); Joshua B. Fischman & Max M. Schanzenbach, *Racial Disparities Under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums*, 9 J. EMPIRICAL LEGAL STUD. 729, 729 (2012); Johnson & Betsinger, *supra* note 38, at 1079; Max Schanzenbach & Michael L. Yaeger, *Prison Time, Fines, and Federal White-Collar Criminals: The Anatomy of a Racial Disparity*, 96 CRIMINOLOGY 757, 781 (2006) (focusing on white-collar offenders). Still, there is at least one study that concluded there are not disparities based on race/ethnicity when the outcome was operationalized as life sentences. Hamilton, *Life Sentences*, *supra* note 111, at 848 tbl. 4 (finding no statistically significant racial/ethnic differences in long sentences (operationalized as at least 470 months) in federal sentencing in a model with multiple controls).

[116] Cyndy Caravelis et al., *Static and Dynamic Indicators of Minority Threat in Sentencing Outcomes: A Multi-Level Analysis*, 27 J. QUANTITATIVE CRIMINOLOGY 405, 407 (2011).

[117] *E.g.*, Tillyer et al., *Gender*, supra note 108, at 713 tbl. 2 (citing studies and reporting on study of drug offenses); RHODES ET AL., supra note 37, at 67; David B. Mustard, *Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts*, 44 J.L. & ECON. 285, 300 (2001).

chivalry thesis in which paternalistic ideologies conceive of women in ways that reduce their blameworthiness, such as perceiving females as more childlike, less responsible for their own behavior, in need of male protection, and whose suffering should be kept to a minimum.[118] In addition, it might be relevant to judges that women consistently show at lower risk of recidivism.[119]

In some studies, noncitizens are at a statistically significant greater likelihood of incarceration[120] and an increase in sentence length compared to citizens.[121] A theory for why noncitizenship might lead to more punitive outcomes is that persons presenting with an attribute that makes them culturally dissimilar to the American-born population might be adjudged more negatively as outsiders and thereby subject to marginalization in a socially stratified society.[122] Still, an opposing theory argues persons not legally resident in the United States are deportable and thus a longer sentence may be unnecessary.[123]

Studies commonly indicate that older offenders are treated more leniently than their younger counterparts.[124] It could be the negative correlation between older age and severity of penalty is not just about age per se, but a combination of age, infirmity, and physical impairment may lead to an empathetic response.[125] The impact of age may also be for the focal concern of future dangerousness as older offenders are less likely to recidivate.[126]

Two case-processing factors are relevant to predicting sentencing decisions. The so-called trial penalty occurs when being found guilty at trial (rather than plead) is correlated with more serious punishments.[127] The trial penalty may be

---

[118] S. Fernando Rodriguez et al., *Gender Differences in Criminal Sentencing*, 87 Soc. Sci. Q. 318, 320 (2006).

[119] *See generally* Tonya L. Nicholls et al., *Female Offenders, in* APA Handbook of Forensic Psychology 79 (Brian L. Cutler & Patricia A. Zapf eds., 2nd ed., 2015) (reviewing studies and rationales for females being less risky).

[120] Michael T. Light, *The New Face of Legal Inequality: Noncitizens and the Long-Term Trends in Sentencing Disparities Across U.S. District Courts, 1992-2009*, 48 Law & Soc'y Rev. 447, 464 tbl. 2 (2014) [hereinafter Light, *Noncitizens*]; Johnson & Betsinger, *supra* note 38, at 1067 tbl. 3.

[121] *E.g.*, Kim et al., *supra* note 43, at 1084 tbl. 2; Light, *Noncitizens*, *supra* note 120, at 466; Mustard, *supra* note 117, at 301. Though, at least one other study are to the contrary, showing that lacking citizenship has a suppressing impact on the length of the term of imprisonment. Oleson et al., *supra* note 11, at 323 tbl. 2, 325 tbl. 3 (though the statistic was not statistically significant).

[122] Light, *Noncitizens*, *supra* note 120, at 455.

[123] Scott E. Wolfe et al., *Unraveling the Effect of Offender Citizenship Status on Federal Sentencing Outcomes*, 40 Soc. Sci. Res. 349, 352 (2011).

[124] *E.g.*, Anita N. Blowers & Jill K. Doerner, *Sentencing Outcomes of the Older Prison Population: An Exploration of the Age Leniency Argument*, 38 J. Crime & Just. 1, 3-4 (2013) (citing studies); Johnson et al., *Social Context*, *supra* note 52, at 761 tbl. 5 (finding older age positively correlated with downward departure decisions); John D. Burrow & Barbara A. Koons-Witt, *Elderly Status, Extraordinary Physical Impairments and Intercircuit Variation Under the Federal Sentencing Guidelines*, 11 Elder L.J. 273, 312-13 tbl.3, 4 (2004) (finding that in a few districts defendants age 50 and over were more likely to receive downward departures).

[125] Burrow & Koons-Witt, *supra* note 124, at 296.

[126] Franklin, *Educational Attainment*, *supra* note 38, at 142.

[127] *E.g.*, Kim et al., *supra* note 43, at 1084 tbl. 2; Andrew Chongseh Kim, *Underestimating the Trial Penalty: An Empirical Analysis of the Federal Trial Penalty and Critique of the*

about punishing those who have the "temerity to go to trial."[128] It could be viewed instead in terms of rewarding pleas, such as rewarding cooperation and remorse while also preserving court resources.[129]

As for the second case-processing factor, studies at the state and federal levels rather consistently show that pretrial detention is significantly and positively related to incarceration and sentence length.[130] Pretrial detention effects are likely due to the same drivers as the focal concerns perspective posit. Those who are denied release pretrial may be more likely to have committed a more serious crime, bear a significant criminal history, and present with other indicators that elevate their potential recidivism risk.[131]

Studies which include district or circuit variables in their models have generally found geographic disparity in federal sentences.[132] These outcomes lend support to the court communities' perspective of localized practices influencing case decisions and fostering regional differences in federal sentencing.

2. The Outcome of Interest in Prior Studies

A significant majority of the foregoing studies on federal sentencing use the incarceration decision (in/out) and/or sentence length as their outcome of interest. Some researchers affirmatively, though, recognize the importance of investigating departure decisions. Almost all of the studies of federal departure decisions to date which model the dependent variable on departure outcomes address downward departures.[133] Decisions to depart downward are certainly deserving of study because a significant percentage of federal sentences these days are below their Guidelines minimums.[134] None of the previous empirical studies appear to have focused extensively on the effect of upward departures as the outcome of interest.

Abrams Study, 84 Miss. L.J. 1195, 1220 (2015); Breen, *supra* note 101, at 211 (citing studies).

[128] Michael M. O'Hear, *Plea Bargaining and Procedural Justice*, 42 Ga. L. Rev. 407, 409 (2008).

[129] Ulmer et al., *Trial Penalties, supra* note 108, at 564.

[130] *E.g.*, Oleson et al., *supra* note 11, at 316-17 (citing studies); Franklin, *Educational Attainment*, *supra* note 38, at 151 tbl. 2; Wolfe et al., *supra* note 123, at 355 tbl. 2. There is one study to the contrary where being out on bail increased sentence length, which the authors did not expect and do not explain the result. Farrell et al., *supra* note 113, at 115 tbl. 3.

[131] Oleson et al., *supra* note 11, at 317. Correspondingly, a judge may perceive a defendant who is released on bail and complies with supervision as presenting with a positive rehabilitation potential. *Id*.

[132] *E.g.*, Wu & Spohn, *supra* note 52, at 306 (finding differences in the likelihood of downward departures across three Midwestern districts); Ulmer, *Localized Uses*, *supra* note 63, at 269 (finding from a study of four districts significant variations in the likelihood of granting substantial assistance downward departures).

[133] *E.g.*, Tillyer & Hartley, *supra* note 7, at 635 tbl. 2; Kimberly A. Kaiser & Cassia Spohn, *Fundamentally Flawed? Exploring the Use of Policy Disagreements in Judicial Downward Departures for Child Pornography Sentences*, 13 Criminology & Pub. Pol'y 241 (2014); Melissa A. Logue, *Downward Departures in US Federal Courts: Do Family Ties, Sex, and Race/Ethnicity Matter?*, 34 Ethnic & Racial Stud. 683 (2011); Johnson et al., *Social Context*, *supra* note 52.

[134] U.S. Sent'g Comm'n, 2015 Sourcebook tbl. N (2016).

This is curiously true, despite upward departures arguably being more substantial, such as leading to longer sentences in the face of the federal prison overpopulation. Plus, their relative rarity renders upward departures more symbolic in nature, perhaps perceived therefore as arbitrary. Almost all the studies to date which consider the upward departure decision as a variable at all simply add it as a control without further discussion of its significance because their interests concerned other aspects of sentencing.[135]

It appears that only three studies (two of them by the same author) have so far utilized the upward departure decision as an outcome variable. Nevertheless, in these trio of studies the upward departure decision was one of multiple outcomes in single-level regressions and the authors did not spend too much space delving into the upward departure's importance in federal sentencing outcomes.[136] The earliest study utilized pre-*Booker* data and controlled only for sociodemographic characteristics.[137] The researcher's attention in the other two studies concerned *Booker*-based variations in sentencing outcomes more generally and the potential, more specifically, for courtroom disparities before and after *Booker* (finding greater disparity in upward departures post-*Booker*)[138] and racial disparities (finding greater racial disparities in upward departure decisions post-*Booker*).[139] This latter author in one study tested a subset of the Commission's data for the time period of study[140] and reports little in either paper of the effects of explanatory factors tested with respect to upward departures (other than race and the *Booker* time trend) and for some reason excluded many predictor variables found to be relevant to sentencing outcomes.[141]

Due to the paucity of research with a concentration on the upward departure

[135] *E.g.*, Franklin et al., *Intermediate Sanctions*, *supra* note 5, at 870 n.12; Tillyer et al., *Gender*, *supra* note 108, at 713 tbl. 2; Ulmer & Light, *supra* note 39, at 336 tbl. 2; Feldmeyer & Ulmer, *supra* note 39, at 252 tbl. 3.

[136] Crystal S. Yang, *Free at Last? Judicial Discretion and Racial Disparities in Federal Sentencing*, 44 J. Leg. Stud. 75, 95-98 (2015) [hereinafter Yang, *Discretion*]; Crystal S. Yang, *Have Interjudge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence from* Booker, 89 N.Y.U. L. Rev. 1268, 1314 (2014) [hereinafter Yang, *Interjudge*]; Mustard, *supra* note 117, at 305-09.

[137] Mustard, *supra* note 117, at 310 tbl. 11.

[138] Yang, *Interjudge*, *supra* note 132, at 1315.

[139] Yang, *Discretion*, *supra* note 136, at 98.

[140] Yang, *Interjudge*, *supra* note 132, at 1296.

[141] Yang, *Discretion*, *supra* note 136, at 98 (indicating in model for upward departures included only predictor variables regarding race, time frame based on United States Supreme Court rulings such as *Booker*, offense type, offense level, criminal history, district, year, and month); Yang, *Interjudge*, *supra* note 132, at 1314-15 (controls included time variables, offense type, offense level, criminal history, and districts). In the 2015 report, the author's conclusion with a logistic regression analysis was that for fiscal years 1994-2010 blacks were more likely (with statistical significance) to be assigned upward departures than whites. I was generally able to replicate this result using the Commission's full dataset for most of the time period of study (fiscal 1999-2010) following the paper's indication of methodology and control variables except for the *Booker* timing and sentence month. However, by re-specifying the model with additional, statistically significant controls, the coefficient for blacks (compared to whites) became nonsignificant. This means that the difference indicated for racial disparity in the other researcher's model appears to be explained away by the addition of other legal and extra-legal factors (specifically, the variables I added were acceptance of responsibility, custody status, number of counts, gender, citizenship, and age).

decision, the importance of it in the results of sentencing outcomes in terms of severity of sentence, and the symbolic nature of the discretionary decision with respect to potentially reflecting gaps in the Guidelines, the opportunity to fill the void was compelling. Then the recent availability of more aggressive computing resources to permit employing a sophisticated research design known as multilevel modeling would allow this study to also be able to test for possible regional disparities. Hence, the next Section offers such a study.

IV. A MULTILEVEL STUDY OF UPWARD DEPARTURES

The most common type of advanced statistical analysis of sentencing outcomes is a single-level regression model with individual predictors.[142] At its simplest, a regression can test the relationship between an independent (also known as predictor or explanatory) variable and the dependent (also referred to as outcome or response) variable of interest.[143] It is unlikely, though, for any outcome of interest in the complex world of criminal justice to be fully explained by one independent factor.[144] Certainly, the focal concerns and courtroom workgroup perspectives would predict that numerous factors would play a role in individual criminal justice outcomes. Helpfully, sophisticated regression models permit a researcher to test the effects of a host of independent variables on the chosen dependent variable, and most current regression studies appropriately utilize multiple predictors. A value of a multiple regression analysis is that a researcher can investigate the effect of each independent variable on the dependent variable while controlling for (i.e., holding constant) the effect of other explanatory variables.[145] For example, if the researcher is interested in whether race is associated with sentence length, she likely ought to include offense severity and criminal history (at the very least) in the model to control for them as it could be that the association between race and sentence length may be largely explained by such legal factors.

Sentencing research now seems on the precipice to replacing single-level regressions with the more sophisticated technique of multilevel modeling.

A. MULTILEVEL MODELING

The concept of multilevel modeling is a relatively recent development in the field of statistics.[146] The growth of interest in conducting multilevel modeling in the last decade is likely based on several factors. Some researchers have realized the

---

[142] Cassia Spohn, *The Evolution of Sentencing Research*, 14 CRIMINOLOGY & PUB. POL'Y 1, 2 (2015).

[143] RONET BACHMAN & RAYMOND PATERNOSTER, STATISTICAL METHODS FOR CRIMINOLOGY AND CRIMINAL JUSTICE 489 (1997).

[144] *Id*.

[145] Paul Hofer, *The Commission Defends an Ailing Hypothesis: Does Judicial Discretion Increase Demographic Disparity?*, 25 FED. SENT'G REP. 311, 311 (2013).

[146] ANTHONY S. BRYK & STEPHEN W. RAUDENBUSH, HIERARCHICAL LINEAR MODELS: APPLICATION AND DATA ANALYSIS METHODS 3-4 (1992).

flaws in single-level designs when the units of analysis are nested within groups where group-level factors affect the outcome of interest.[147] As a result of this early research, knowledge about multilevel models is starting to become more readily available in scientific literature.[148] In addition, technological improvements in statistical software and hardware computing ability make the resource-intensive analysis of multilevel data more accessible and workable.[149]

In discussing multilevel models, the terminology typically entails levels, usually in a linear fashion to signify the nesting structure. Level-1 is the most elemental. Level-1 units are clustered at Level-2. Three-level models involve Level-2 clusters that are nested into a higher order. For instance, as visually represented in Figure 1, federal sentencing entails a hierarchical structure in which individual defendants represent Level-1 units, with district courts at Level-2, and circuit courts representing Level-3.

Multilevel methods permit the researcher to specify an explanatory variable as a fixed effect, a random effect, or both. A fixed effect variable specifies a single value in the model and is applicable to each Level-1 unit, regardless of which Level-2 group the unit is situated.[150] The coefficient of a fixed effect variable acts like an explanatory variable in a single-level regression analysis, indicating the variable's effect on the outcome of interest. In the study herein, individual defendants comprise Level-1, such that the fixed effects test for how the unique attributes of the individual defendant impacts whether an upward departure is issued. As an example, the study tests whether the defendant's gender is correlated with an upward departure.

A random effect, on the other hand, allows an explanatory variable to vary between Level-2 units such that each Level-2 group has its own estimate of that variable.[151] It should be noted that a random effect does not signify that it is unsystematic, occurs by chance, or is unexplained. Instead, a variable being specified as random refers to observing whether its effect on the dependent variable fluctuates over Level-2 groupings.[152] For our purposes in this paper, a random effect tests whether, for example, even if gender is found overall to be a significant individual predictor of an upward departure, the same effect is consistently observed (or not) across district courts.

A random effect coefficient for a predictor variable that is statistically significant, for purposes of the study herein, indicates that (a) the magnitude (i.e., strength) of the effect of the variable is weaker in some districts but stronger in other districts, and possibly (b) that the effect of that variable changes direction across districts

---

[147] Brian D. Johnson & Christina D. Stewart, *Measurement Issues in Criminal Case Processing and Court Decision-Making Research, in* The Handbook of Measurement Issues in Criminology and Criminal Justice 303, 314-15 (Beth M. Huebner & Timothy S. Bynum eds., 2016) (citing multilevel modeling research sources in criminal justice).

[148] *E.g., see generally* Joop J. Hox, Multilevel Analysis: Techniques and Applications (2nd ed., 2010); Leonardo Grilli & Carla Rampichini, *Specification of Random Effects in Multilevel Models: A Review*, 49 Qualitative & Quantitative 967 (2015).
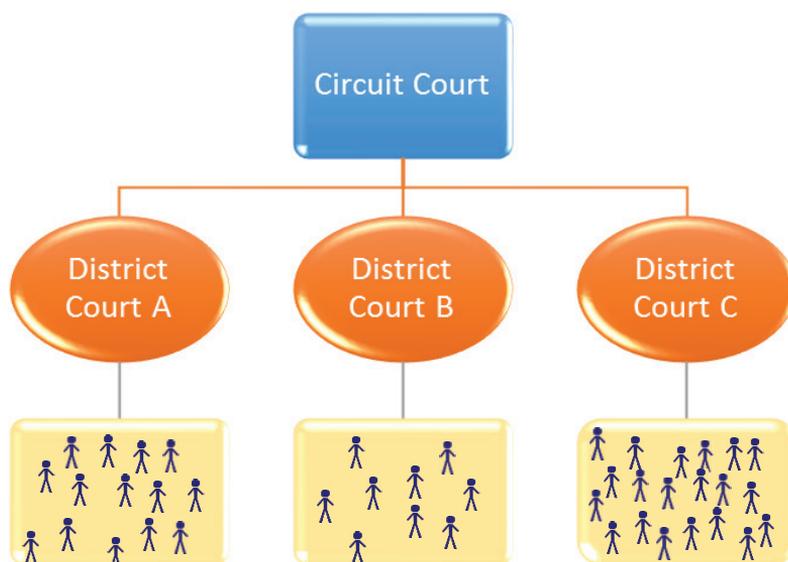
[149] Daniel A. Powers, *Multilevel Models for Binary Data*, 154 New Directions Inst. Res. 57, 62 (2012).

[150] Andrew F. Hayes, *Multilevel Modeling*, 32 Human Comm. Res. 385, 389 (2006).

[151] *Id*.

[152] Tom A.B. Snijders, *Fixed and Random Effects, in* Encyclopedia of Statistics in Behavioral Science 664, 664 (Brian S. Everitt & David C. Howell eds., 2005).

Figure 1. Example of a Three-Level Model for Federal Sentencing.

units from positive to negative, or vice versa.[153] As an hypothesized example of (b), it could be that criminal history is a positive predictor in some districts, meaning that the higher criminal history score increases the likelihood of an upward departure; yet, criminal history could be a negative predictor in other districts, such that a higher criminal history score decreases the chance of an upward departure. A random effect that is not statistically significant may still provide meaningful information. A non-statistically significant random effect indicates that the effect of that predictor variable on the outcome fails to differ across districts such that the effect is not group-dependent (here, this means the relationship between the predictor and an upward departure is relatively consistent across districts).

A multilevel study that includes both fixed and random effects is generally referred to as a mixed model. One of the strengths of specifying multilevel modeling is the ability to test whether a particular explanatory variable may have different effects at each level. An explanatory variable may be statistically significant at Level-1 (the fixed effect) and may—or may not—show statistical significance at Level-2 (the random effect), or vice versa.[154]

Overall, multilevel modeling presents an advancement for statistical research in criminal justice. In regards to penalty outcomes, it is particularly important to focus on both (a) individual level predictors because of the focal concerns perspective, and (b) on jurisdictional level variations because there may be relevant contextual

---

[153] John Wooldredge, *Judges' Unequal Contributions to Extralegal Disparities in Imprisonment*, 48 CRIMINOLOGY 539, 549 (2010).

[154] Multilevel modeling can thereby overcome aggregation bias that exists when an explanatory variable shows different results at different levels. BRYK & RAUDENBUSH, *supra* note 146, at 83.

differences stemming from unique cultural characteristics or peculiarities produced through discrete courtroom community practices.[155] Further information on the theoretical, statistical, and practical values of multilevel modeling can be found in the Appendix to this paper.

Despite the many advantages of multilevel modeling techniques, relatively few multilevel studies have been conducted in federal sentencing. This does not mean that many other researchers have not been cognizant of the potential that geographical and jurisdictional differences may have significant impacts on individual sentencing outcomes. Typically, researchers realizing the potential for regional differences in federal sentencing simply control for these group-level variances in single-level regression models by adding districts[156] or circuit courts[157] as a series of dummy variables. It was certainly proper to account for at least some of the variation that district and circuit courts may introduce to sentencing outcomes. Yet these single-level regression models were unable then to take advantage of the benefits of multilevel modeling, and it is possible that at least some of the results in those studies were therefore biased.

The rather scant number of studies which do apply a better specified model from a methodological perspective by adapting multilevel modeling to federal sentencing data have tended to focus on sentence length as the outcome of interest.[158] Several researchers have studied departure decisions in multilevel designs, though they concentrate on downward departures as the dependent variable.[159] In any event, these studies typically utilized pre-*Booker* data[160] and, therefore, may no longer be generalizable to the current state of affairs. This study supplements the existing literature by addressing upward departures, drawing upon a lengthy period of post-*Booker* sentencing practices, and providing a mixed model with a host of fixed and random effect explanatory variables. The data and methods are next summarized.

B. Data and Methods

This study used Commission datasets for the fiscal years 2008-2015 to represent a long period of sentencing practices and to account for post-*Booker* discretionary decisionmaking. These datasets offer a host of variables parsing individual sentence details. The Commission codes the variables based on a variety of documents: the

---

[155] Gaylene S. Armstrong & Nancy Rodriguez, *Effects of Individual and Contextual Characteristics on Preadjudication Detention of Juvenile Delinquents*, 22 Just. Q. 521, 525 (2005).

[156] *E.g.*, Franklin et al., *Intermediate Sanctions*, *supra* note 5, at 860 tbl. 4; Joshua B. Fischman & Max M. Schanzenbach, *Racial Disparities Under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums*, 9 J. Empirical Legal Stud. 729, 740 (2012); Ulmer & Light, *supra* note 39, at 334; Johnson & Betsinger, *supra* note 38, at 1068 tbl. 3; Mustard, *supra* note 117, at 300.

[157] *E.g.*, Blowers & Doerner, *supra* note 124, at 8; Doerner & Demuth, *supra* note 110, at 254; Melissa Hamilton, *Sentencing Policy Adjudication and Empiricism*, 30 Ga. St. U.L. Rev. 375, 454 tbl. 3 (2014).

[158] Ulmer et al., *Trial Penalties, supra* note 108, at 575; Lynch & Omori, *supra* note 104, at 423.

[159] Tillyer & Hartley, *supra* note 7, at 1631; Johnson, Ulmer, & Kramer, *supra* note 52, at 750.

[160] *E.g.*, Feldmeyer & Ulmer, *supra* note 39; Albonetti & Baller, *supra* note 39; Farrell et al., *supra* note 113, at 103; Kautt, *supra* note 90, at 648.

judgment and commitment order, the Statement of Reasons, any plea agreement, the indictment, and the presentence investigation report.[161]

There are three main research questions:

1. Is there significant variation across district courts in the use of upward departures?
2. To what extent do legal, extralegal, and case-processing factors account for upward departures in individual cases?
3. Do district courts vary from each other in the extent to which they weigh each of the legal, extralegal, and case-processing factors when issuing upward departures?

In the multilevel design, the outcome (dependent) variable is whether the judge issued a sentence that was an upward departure from the Guidelines recommendation. This outcome and a list of the multiple predictor variables (comprising legal, extralegal, and case-processing factors) which survived to the final multilevel model and their coding are summarized in Table 1.

In addition to the multilevel models, a statistical analysis was conducted concerning just the upward departure cases. Commission rules direct district judges when departing from the Guidelines to state the reasons for the departure and to specifically record them in the Commission-generated Statement of Reasons form that is submitted with the paperwork for each individual sentencing.[162] These

Table 1. Coding Scheme of Variables.

Variable	Coding Scheme	Description
Dependent Variable		
Upward Departure	1 = yes	Defendant received an upward departure
Predictor Variables		
Legal		
Final Offense Level	Scale	Guidelines scale rating offense severity from 1-43
Criminal History	Ordinal	Guidelines ranking of criminal history from I-VI
Number of Counts	Log (scale)	Natural log of the number of counts of conviction
General Offense Type	Five dummy variables	Five dummy indicators with the reference category of drug offenses
Acceptance of Responsibility	1 = yes	Dummy indicator for having received a reduction in offense levels for accepting responsibility
Extralegal		
Male	1 = male	Dummy indicator for gender
Minority	1 = minority	Dummy indicator for black, Hispanic, or other together coded as 1, with the reference category white
U.S. Citizen	1 = citizen	Dummy indicator for a U.S. citizen
Age Over 50	1 = yes	Dummy indicator for age 50 and above
Case-Processing		
In Custody	1 = yes	Dummy indicator for being in custody at time of sentencing
Trial	1 = yes	Dummy indicator for going to trial (versus a plea)
Level-2	Nominal	94 districts

---

[161] CHRISTINE KITCHENS, FEDERAL SENTENCING DATA AND ANALYSIS ISSUES 1 (2010).
[162] U.S. SENTENCING GUIDELINE MANUAL § 5K2.0(3).

are then coded by staff into the Commission's datasets. Thus, a separate analysis (external to the multilevel model) ran frequency distributions of the multiple variables representing the reasons judges provided for the upward departure cases over fiscal years 2008-2015. The results of the multilevel studies and these frequency distributions are provided next.

C. RESULTS

The research questions posed earlier indicated a two-level design with district courts at Level-2. Descriptive statistics regarding the variables that survived to the resulting full model are provided in Table 2.

Separate statistical analyses of Commission datasets (fiscal 2008-2015) indicated that an upward departure is typically of significant consequence to the receiving defendant's sentence: the mean sentence for those defendants receiving an upward departure for the period of study was 84.44 months (about 7 years), with a range from probation to 4,253 months (about 354 years).[163]

Table 2. Descriptive Statistics.

Variable	Mean (%)
Dependent Variable	
Upward Departure	(2.0%)
Predictor Variables	
Legal	
Final Offense Level	18.72
Criminal History	2.48
Number of Counts	1.42
General Offense Type	
Drugs	(33.0%)
Violent	(5.9%)
Firearms	(10.6%)
Immigration	(29.9%)
Property	(16.5%)
Other	(14.0%)
Acceptance of Responsibility	(94.8%)
Extralegal	
Female	(12.8%)
Minority	(73.5%)
U.S. Citizen	(58.7%)
Age Over 50	(12.5%)
Case-Processing	
In Custody	(75.3%)
Trial	(3.5%)

---

[163] The reader may wonder if the 354 year figure is a typographical error or a data error. It is not. This extreme sentence was handed to Corey Deyon Duffey in 2010 for a series of bank robberies. Two of Duffey's co-defendants received similar sentences of 355 and 330 years. Perhaps not surprisingly, the district that sentenced them to these extreme sentences was the Northern District of Texas, the same district that has the highest rate of upward departures in the study period (2008-2015). For more information on the use of extreme sentences such as Duffey's, see generally Hamilton, *Extreme Sentences*, *supra* note 50.

Table 3. Full Multilevel Model of Upward Departures.

Variable	Fixed Effect			Random Effect	
	b	S.E.	Odds Ratio	s^2	S.E.
Intercept	-5.021	.152	-----***	.064	.051
Legal Factors					
Final Offense Level	-.072	.004	.931***	.001***	.000
Criminal History	.057	.013	1.059***	.009***	.002
Number of Counts (log)	.315	.018	1.370***	.009**	.003
General Offense Type				---	---
Drugs (reference)					
Violent	1.576	.116	4.838***		
Firearms	.694	.094	2.001***		
Immigration	.199	.106	1.221		
Property	.532	.096	1.702***		
Other	.503	.116	1.653***		
Acc. of Responsibility	-.728	.070	.483***	.045*	.018
Extralegal Factors					
Female	-.559	.047	.572***	.018	.014
Minority	.045	.044	1.046	.035***	.010
U.S. Citizen	.509	.066	1.664***	.148***	.031
Age Over 50	.311	.031	1.364***	.010	.006
Case-Processing					
In Custody	1.403	.055	4.066***	.055***	.016
Trial	-.100	.084	.905	.063*	.027
Random intercept				.064	.051
ρ				1.9%	
-2LL = 4149605					
n = 567,294					

*p < .05; **p < .01; ***p < .001

The final multilevel model included 567,294 cases and is provided in Table 3.[164] All variables were estimated with both fixed and random effects except for one. The general offense type series of five dummy variables was excluded from random effects for statistical resource reasons, as explained in the Appendix. In Table 3, the left column lists the predictor variables. The middle column indicates their coefficients, standard errors, and odds ratios for the fixed effects. The right hand column lists the coefficients and standard errors for the random effects.

The final model includes a substantial portion of the explanations for upward departures. Overall, the model poses a 98% correct classification rate. This section textually delineates the substantive results, with further discussion to follow in the next Section to explore how the theoretical background regarding focal concerns and the community workgroup thesis may help explain these results.

---

[164] Eleven percent of the potential cases were excluded because of missing data on any one of the final predictor factors. There is no reason to believe the missing cases represent any bias.

1. Individual Disparities

The results for the fixed effects (i.e., individual defendant predictors) will be addressed first. All of the legal factors achieved statistical significance in their individual effects on upward departures. The final offense level was negatively associated with the odds of an upward departure: the odds of an upward departure decreased 7% for every one level increase in the final offense level. The criminal history score had the opposite effect in being positively associated with an upward departure: the odds of an upward departure increased 6% for each one unit increase in criminal history category. The presence of multiple counts of conviction were associated with increased odds of an upward departure. Regarding crime type, compared to drug offenders as the reference category, the other offense types were more likely to receive upward departures. Violent offenders faced almost five times the odds of an upward departure while the odds for firearm offenders doubled. Only immigration offenses did not result in statistical significance. Acceptance of responsibility lowered the odds of an upward departure by a factor of two.[165]

Demographic variables were also modeled as fixed effects. Females were significantly less likely to receive upward departures than males, even after controlling for multiple factors: an upward departure for males was almost two times the odds as for females. U.S. citizens were more likely to be assigned upward departures, with the odds of citizens receiving upward departures being 66% greater as compared to noncitizens. There was also an age effect, with those age 50 and over being more likely to receive an upward departure compared to their younger counterparts.

Minorities were at higher risk of upward departures. The odds of a minority defendant receiving an upward departure increased 5% when controlling for the other legal and nonlegal variables. However, the result at the individual case level (Level-1) for the minority variable was not statistically significant. Still, as will be addressed further below, the minority factor was retained as there was a statistically significant random effect (districts at Level-2) for it, indicating that the lack of significance at the individual case level does not mean there is not a minority effect on increasing the odds of an upward departure in at least some districts.

Both case-processing factors were statistically significant. Custody status exhibited a large effect, increasing the odds of an upward departure by a factor of four for those in custody at sentencing. The trial penalty was not statistically significant at the individual level. However, the trial versus plea factor was retained because, as also addressed below, the random effect coefficient for the trial penalty at the district level indicated statistical significance, signifying that there are trial penalties in at least some districts.

2. District Disparities

The random effects (i.e., variations among districts) of the variables in the far right columns of Table 3 indicate whether the effect of each predictor varied across districts (except offense type which was excluded for statistical reasons per the Appendix). All but two of the predictor factors with random effects (being gender

---

[165] As the coefficient is less than 1.00, we can interpret the effect on the odds by taking the reciprocal of the odds ratio = 1/.483.

and age over 50) were found to vary across districts to a statistically significant degree.

Further information on the variability of each predictor factor that was modeled with fixed and random effects can be provided. Computations adding and subtracting one and two standard deviations indicated by each predictor variable's random effect from the same variable's fixed effect coefficient show whether the variability between districts concerns the strength of the correlation with the outcome and if the direction of the correction is positive in some districts yet negative in others.[166] In other words, a particular variable may have a stronger effect on the upward departure decision in different districts compared to others. The same variable may also have inconsistent effects in that it is predictive of an upward departure in some districts yet is predictive of no upward departure in others.

For six of the random effects, the size of the effect across two standard deviations varied between districts (i.e., across 95% of the districts), but not the direction. The number of counts of conviction, age over 50, and being in custody at sentencing were each positively correlated with upward departures in at least 95% of districts. The final offense level, acceptance of responsibility, and being female were negative predictors of upward departures in at least 95% of districts.

In contrast, the effect of each of criminal history score, minority status, and trial penalty showed that the strength and the direction of its influence changed across just one standard deviation (i.e., two-thirds of districts). This means that not only the size of the effect of these three variables varied amongst districts but that each held a positive effect in at least some districts while indicating a negative impact in others. U.S. citizenship held a positive association with upward departures in one standard deviation, but across two standard deviations the effect was observed to be negative in at least a few districts.

A supplemental data analysis provides further information about the reasons for upward departure decisions derived from the judges' Statement of Reasons forms filed with sentencing paperwork in individual cases. Table 4 contains the top ten cited reasons for upward departures capture through frequency analyses of the Commission's data, along with their prevalence.

Importantly, considering the title of this Article, unwarranted disparities in upward departures as an external consequence was among the top ten rationales as observed in Table 4. Judges cited disparity issues in one out of twelve upward departure decisions. This result indicates that numerous judges remain cognizant of the potential downsides of the appearance of disparities in sentencing practices. It is also suggestive of gaps in the Guidelines to the extent these judges perceive that the Guidelines calculations in the instant cases failed to achieve proportionality with sentences for similarly-situated defendants. The other reasons judges gave as indicated in Table 4 as justifications for upward departures will be explored further in the context of the general discussion of the results that follows.

D. Discussion

The results just provided can now be more fully addressed concerning the three research questions previously posed. Further, they can be better understood in

---

[166] *See generally* Joop, *supra* note 148, at 19.

the context of the theoretical perspectives offered implicating the focal concerns perspective and the courtroom workgroup thesis.

Table 4. Specific Reasons Given by Judges for Upward Departures.

Rank	Reason	Percentage of Cases
1	Criminal history issues	60.0%
2	Nature and circumstances of the offense and history and character of the defendant	53.5%
3	Reflect the seriousness of the offense, promote respect for the law, and provide just punishment	49.9%
4	Deterrence	42.6%
5	Protect the public from further crimes of the defendant	40.9%
6	Rehabilitation	9.3%
7	Avoid unwarranted disparities	8.0%
8	Dismissed and acquitted conduct	8.4%
9	General adequacy issue	5.5%
10	General guideline issue	4.4%

1. Distract Disparities Overall

The first research question queried whether there existed significant variation between district courts in the use of upward departures. The answer is in the affirmative. Bivariate results that were the result of additional statistical analyses indicated a differential of twelve times the rate of upward departures between the lowest rate district and the highest. Significant variation was confirmed in a null multilevel model (see the Appendix) which indicated that 8% of the total variance in upward departure outcomes is explained at the district court level. This rate was statistically significant at the .001 level. In other words, this means that eight percent of the differences in upward departure decisions are accounted for by district court practices. This result of district differences was expected from the courtroom workgroup perspective in that cultures unique to certain districts may influence sentencing outcomes that contrast with outcomes from other cultures/ districts.

2. Individual Disparities

The second general research question asked to what extent legal, extralegal, and case-processing factors accounted for upward departures in individual cases. Generally the results support the influence of the focal concerns (concerning the defendant's culpability and future risk and the consequences of the sentence) on individual outcomes with respect to upward departures.

The legal variables supported the focal concerns expectation that perceptions of the defendant's blameworthiness are highly relevant to individual penalties. The results indicated an increased likelihood of an upward departure for a higher criminal history score, multiple counts of conviction, and violent and firearms

offenses (compared to drug offenders). Criminal history and additional counts signify multiple crimes and perhaps perpetrated on multiple occasions, possibly demonstrating greater culpability and harm. The increased odds for violent and firearms offenses reveal culpability concerns in that crimes posing a risk to human life likely are considered more egregious than many nonviolent offences.

The decreased likelihood of an upward departure for acceptance of responsibility is also consistent with a concern for the defendant's blameworthiness as well as with the focal concern of future risk. Accepting responsibility by admitting guilt at an early stage in the proceeding may be perceived to reduce one's culpability while predicting positive rehabilitation potential. The negative correlation of acceptance of responsibility with upward departures was consistent across at least 95% of districts.

Curiously, the final offense level was negatively correlated with the upward departure decision. This result seems to be somewhat contradictory to the focal concern with greater offender culpability predicting more severe sentences. It may instead, then, suggest that in these cases judges find the Guidelines calculations to be more than sufficiently proportional to reasonable sentences as adjudging offense severity. This explanation is likely because stakeholders tend to find Guidelines recommendations are overly punitive as a general rule.[167]

Further discussion of criminal history is warranted as it played a strong role throughout the results. There were multiple indications that judges perceive inadequacies in the criminal history calculations. As previously indicated, a higher Guidelines-calculated criminal history score increased the odds of an upward departure despite multiple controls. This result implies that judges in these cases do not believe the criminal history calculation is sufficiently proportional to prior offending evidence, at least when the defendant already has a substantial criminal history as officially calculated pursuant to Guidelines rules. This observation is buttressed by the reasons judges listed in explaining upward departures. In the list of rationales judges gave for upward departures from the frequency distributions provided in Table 4, the role of criminal background is salient. Criminal history calculation issues were expressly cited in 60% of the cases, earning the top ranked reason overall for upward departures. Relatedly, as a separately coded reason, evidence of dismissed and acquitted conduct was listed as an explanation for upwardly departing in 8% of upward departures. Further, past offending may be part of the second ranked reason, which includes the history and character of the defendant, cited in over half of the upward departures. Because of the broad nature of that particular reason as including the nature and circumstances of the offense, though, it is difficult to parse what portion of the fifty percent was for prior offending specifically. Still, the failure of the formal criminal history calculation to adequately account for prior offending was evident in a significant majority of upward departures.[168]

---

[167] *See* resources cited *supra* note 43.

[168] It is of particular note that judges candidly admitted the role of dismissed and acquitted conduct in their decisions to upwardly depart in one out of 12 (8%) cases. This finding might be of concern to critics of the real offense system in which individuals are penalized for conduct that is not the subject of conviction. Critics may be even more offended by increases in punishment for acquitted conduct. Here, it is not possible to tell exactly what percentage of those cases represented acquitted conduct, but it is likely that acquitted conduct played a role in at least some of them. These 8% of upward departure cases

Overall, the salience of criminal history is theoretically important for another reason. The function of the defendant's criminal history in the various results implicates the focal concern regarding the defendant's future risk. The inclusion of criminal history in the Guidelines as a principal factor in the recommended sentence is often viewed as the Commission's proxy to adjudge dangerousness.[169]

Regarding future risk as a focal concern, other reasons in Table 4 more directly address dangerousness. The inclusion of the character of the defendant within the second ranked reason may well include assessments of past antisocial behavior as reflective of future risk. Ranked fifth in the top reasons given, the need to protect the public, clearly a future risk rationale, represented 41% of the upward departures. In sum, the relevance of the focal concern of future risk to severity in sentences is strongly confirmed in the data.

The multilevel results concerning offense type likewise provide interesting information about compliance with Guidelines' proportionality judgments. The dummy series for offense type indicated that all other offense types, except for immigration offenses, were more likely to receive upwards departures than drug cases as the comparator. This implies that district judges as a general rule tend to believe the Guidelines are sufficiently punitive for drug offenses and immigration offenses. As drug and immigration cases combined are the bulk of federal sentencing in percentage terms, this particularly result situate the Guidelines in a positive light in terms of proportionality, at least with respect to generally being sufficiently punitive for a majority of crimes. However, the greater likelihood of upward departures for violent and firearms offenses implies that the judges may perceive the Guidelines as insufficiently punitive in those cases.

Moving onto the impact of extralegal variables, demographic characteristics presented with some expected results, while others were more surprising. There was support for gender leniency as women were far less likely to receive upward departures than men at the individual case level. Plus, gender leniency for women did not vary among districts, even after controlling for a host of other variables. This was the case even though gender is an *extra*legal factor and a prohibited rationale for sentencing outcomes per the Guidelines. Overall, then, the results indicate gender disparities, possibly even gender discrimination in favor of women, in upward departures.

Contrary to many studies, the results here indicate there was no individual-level minority discrimination in upward departure decisions. While the odds for minorities were 5% greater than whites, the result was not statistically significant. Indeed, minority status was the weakest individual predictor overall.[170] A reason that this result is inconsistent with other research finding disparities for minorities may be the greater number of explanatory variables in this model and its ability to

---

may also imply there are instances in which judges are countering plea bargaining to the extent that some percentage of these cases may represent increased penalties due to offenses dismissed as part of plea bargain deals. Perhaps this reflects judges acting as a check on prosecutorial authority in cases in which they view the plea bargains as overly lenient.

[169] Marjorie A. Meyers, Criminal History: Calculation and Variance 1 (2012), available at http://www.ussc.gov/sites/default/files/pdf/training/annual-national-training-seminar/2012/3_Criminal_History-Calculation_and_Variance.pdf (presentation at U.S. Sent'g Comm'n Annual Training Conference).

[170] This result derives from F statistic comparisons.

parse district-level variations. Indeed, the random effect was significant, indicating that minority status matters more in at least some districts. Plus, within one standard deviation, the results indicate there are some districts in which minority status is positively correlated with upward departures, despite numerous controls. Hence, it remains possible that there is explicit or implicit minority discrimination in some regions regarding upward departures, though not throughout the country.

It was surprising that noncitizenship was not a positive predictor of upward departures. Perhaps the explanation for the statistically greater likelihood of United States citizens to receive upward departures is that (according to a supplemental data analysis) two-thirds of the noncitizens in federal sentencing during the period of study (fiscal 2008-2015) were immigration offenders. Noncitizen immigration violators are likely to be subject to deportation. Deportation as an incapacitating gesture may impact an assessment of future risk at least regarding the danger to U.S. residents. Thus, it is possible that for noncitizen immigration offenders, prosecutors typically did not request upward departures in those cases and/or judges may have perceived them as unnecessary because of the deportation option. Still, the random effect of citizenship was statistically significant, indicating that the strength of the effect of citizenship significantly varied between districts. At two standard deviations, the effect of noncitizenship shows that it is actually positive (i.e., noncitizens were at higher odds of upward departures) in at least some districts.

No age leniency was observed at least to the extent it means less punishment for older offenders. Indeed, those age 50 and above were more likely to receive an upward departure and, like gender, the strength of the effect did not vary across districts. This could be evidence of a policy dispute with the Commission's rule that age should typically not be a relevant sentencing factor. An alternative explanation, and one more likely considering the existence of other studies affirming age leniency,[171] relates to the results for criminal history previously discussed. The Guidelines computation of criminal history points contain statute of limitations-types of provisions in which dated offenses are excluded.[172] Simply by virtue of their age, older offenders would be more likely to have offenses far in the past that would be subject to the time bar. In addition, the Guidelines do not count certain types of convictions, such as convictions by military, tribal, and foreign courts and those that resulted in diversion.[173] Older offenders would obviously have a longer opportunity to rack up more convictions by various entities. Altogether, the results strongly indicate that many judges may disagree with such policies for criminal history and thus deviate upward as a result, which would more severely impact older offenders.

In terms of case-processing variables, the failure to find a trial penalty at level-1 is inconsistent with much other research.[174] However, the result here at the individual defendant level is explained by the presence of the acceptance of responsibility variable. Without controlling for the acceptance of responsibility, a previously run multilevel model (with the other predictor variables in Table 3) showed a statistically significant trial penalty factor. Once the acceptance of responsibility variable was input, the significance of the trial penalty vanished.

---

[171] *See supra* note 124 and accompanying text.
[172] U.S. Sentencing Guideline Manual § 4A1.2(d), (3).
[173] *Id*. at § 4A1.2(f)-(i).
[174] *See supra* text and sources accompanying notes 127-129.

Still, the random effects coefficient was significant, and at one standard deviation, the results indicate a trial penalty in at least some districts, which is in line with prior research.

As the last predictor variable to be discussed, custody status was the strongest factor in elevating the odds of an upward departure among the predictor variables.[175] This result affirms that outcomes at sentencing are not entirely independent of decisions at earlier stages in the prosecution process. A denial of pre-trial bail is likely a proxy that influences stronger focal concerns concerning the defendant's culpability for the current offense and greater potential for future dangerousness. Being held in custody through sentencing as a positive predictor of an upward departure was consistent across at least 95% of districts.

The third focal concern should also be mentioned regarding consequences of the penalty. Several of the top reasons judges indicated on the Statement of Reasons for upward departures (listed in Table 4) implicate external consequences. The third highest ranking justification includes respect for the law, which likely entails respect by the defendant individually and more broadly. The fourth reason cites a general deterrence function as a reason for the upward departure, being triggered in 43% of cases. Both reasons reflect upon the consequences of the penalty in its deterring potential offenders and promoting community safety. Another community consequence present among the top ten reasons relates to the rehabilitation of the offender. The frequency of the rehabilitation motive to justify an upward departure, present in 9% of cases, is curious as federal law specifically dictates that "imprisonment is not an appropriate means of promoting correction or rehabilitation."[176] The data do not provide an explanation for the seeming contradiction. Yet it is still relevant as reflecting thoughts toward returning more conforming defendants to their local communities.

Additional evidence exists that upward departure decisions are quite often about proportionality concerns. Rounding out the top ten reasons listed for upward departures are two categories that expressly indicate judicial perceptions that the Guidelines have gaps. Judges cited general guideline issues or general adequacy issues in up to 10% of upward departure cases.

3. District Disparities on Individual Predictors

The third broad research question queried whether district courts vary from each other in the extent to which they weigh each of the legal, extralegal, and case-processing factors when issuing upward departures. The results found numerous such variations, as has already been partly covered when discussing the second research question. Overall, significant random effects were observed for all but two of the predictor variables (excluding offense type which could not be modeled as random effects). The strengths of the effect of leniency for women and the lack of lenience for older offenders were consistent across districts. In contrast, minority status and the trial penalty, which were not statistically significant in individual cases (after controlling for other variables), achieved significance in their random effects. In general, these random effect results support the courtroom communities' perspective which theoretically accounts for different regional sentencing patterns.

---

[175] This result derives from F statistic comparisons.
[176] 18 U.S.C. §3582(a).

To cite two examples, criminal history score and U.S. citizenship were both significant positive predictors of upward departures in individual cases, yet they also held significant random effects, meaning that their relationship to upward departures varied between districts. Moreover, standard deviation computations indicated that criminal history and the citizenship effect were actually negative predictors in some regions.

The discussion shall end on an empirical note. Overall, the results provide strong reinforcement for modeling sentencing decisions with both fixed and random effects in a multilevel model to observe individual- and group-based factors. The statistical significance of multiple explanatory variables in fixed and random effects is itself informative. Then it is also of practical and empirical import that the statistical significance of four variables posed contrasts between their fixed and random effects. In sum, females and age over 50 were statistically significant at their fixed effects, with females and defendants under age 50 far less likely to be issued upward departures (controlling for other explanatory factors). However, there were no significant random effects for those two variables, meaning that the leniency to females and the lack of leniency for those over 50 years-of-age were consistent between districts. The fixed and random effects for two other variables were in the opposite directions. Minority status and going to trial, indicated no significant fixed effects, but their random effects were significant. For minorities and the trial penalty, this means that there are at least a few districts in which minority status is correlated with upward departures and that the trial penalty exists to some extent in at least some districts. The mixed multilevel model employed here was uniquely able to parse those contrasts between individual-level and group-level effects for these four explanatory variables.

V. Conclusions

This Article provided an original empirical study of a discretionary sentencing outcome that leads to more severe sentences. The results show that the focal concerns of culpability, risk, and consequences are significantly relevant to upward departure decisions. Legal and case processing factors regarding these focal concerns are predictive of upward departures and typically in the direction anticipated. The surprising result here was that while higher criminal history score increases the likelihood of an upward departure, the Guidelines offense severity measure produces the opposite effect. A likely explanation is evidence that Guidelines as a general rule offer sufficiently or overly punitive recommendations regarding offense severity. Yet for criminal history, the exclusion of various past crimes in the official Guidelines calculations insufficiently values past antisocial behavior.

It was also of interest that the trial penalty, relevant to culpability and case-processing consequences, is not evident at the individual case level. The explanation is the inclusion of the acceptance of responsibility factor which mediates the trial penalty as a predictor across individual cases. Still, the random effects results also indicate that there exists a trial penalty in at least some districts, even with the acceptance of responsibility variable.

The results confirm that extralegal variables impact non-Guidelines sentences. Leniency for women is strongly supported and systematic, being significant and present across districts. The effect defies the Guidelines policy prohibition consideration of gender. For those who believe gender disparities equal gender discrimination, these results suggest such discriminatory practices. An age effect exists with older age (operationalized as 50 years) being more likely to receive upward departures and, like gender, it was systematically present.

No minority effect is observed at the individual level, though the random effects indicates its presence in at least some districts, even with multiple control variables. Thus, the study finds some racial/ethnic disparities which might constitute implicit or explicit discrimination in some regions. The failure to find that minority status as a consistent predictor of more severe sentences in this study could be due to the multitude of variables measured as fixed and random effects. In turn, citizenship produces an odd result with U.S. citizens more likely to receive upward departures. This result is likely due to the deportation option for non-citizens who commit crimes. On the other hand, this rationale appears to challenge the Guidelines policy that national origin should never be relevant.

Overall, the study suggests reasons for individual disparities in federal sentencing. Likely these embody a mix of warranted and unwarranted disparities, depending upon how one defines and values those terms. The research demonstrates the existence and salience of regional disparities, as well. The multilevel mixed model was able to parse differences between district courts concerning the impact of various legal and extralegal explanatory factors. The results indicate that while gender and age reflect systematic effects, districts vary significantly in their judgment about the relevance of the other predictor factors on upward departure decisions. These variations are consistent with the courtroom workgroup perspective. The results also support the observation that federal courts do not necessarily exhibit a singular culture, share an affinity toward the reasonableness of Guidelines recommendations, or regard national uniformity as the primary goal in sentencing.

This Article contributes to the empirical legal studies literature regarding sentencing practices. It may likewise be helpful more broadly to stakeholders and researchers across criminal justice contexts. The theoretical, policy, and empirical offerings herein may inform about more modernized ways to conceptualize, shape, and study criminal justice outcomes. The study further provides more data in the overall debate about the divergent values of disparity and uniformity.

VI. Methodological Appendix

This Appendix contains additional information about the practical benefits and statistical specifications for multilevel models. It provides the results of several null models (i.e., before explanatory variables were included), further explains some of the independent factors that were transformed in the full model provided in the text of this Article, and discusses why certain other variables were tested yet excluded from the final model.

A. The Limitations of Single-Level Regression Models

Most sophisticated research on sentencing outcomes utilizes single-level regression analysis. While these types of regressions have confirmed values in being able to test the effect of each independent variable in the model while holding constant other variables, there may be an empirical flaw to be recognized in a single-level design as applied to certain datasets. A statistical presumption of a single-level regression model is that the outcomes are independent from one another.[177] Applying this presumption to a study on federal sentencing, like the one presented in this paper, it would mean that a single-level regression model's imperative would be that the impact of, say criminal history score as an example, on the penalty outcome is the same for every defendant, no matter where he or she is sentenced. However, that assumption is likely invalid. Instead, defendants sentenced in the same district court likely share some correlated characteristics. As an illustration, districts at the border of Mexico address a disproportionate percentage of Hispanic defendants committing immigration crimes compared to nonborder districts.[178] The impact of a computed criminal history score on sentences in border districts may vary from other regions simply because border district judges may be aware that official criminal history in foreign countries may not be available in domestic records.[179] Thus, judges facing large numbers of noncitizen defendants may account for the lack of available criminal history information in other ways, thereby skewing the impact of the Guidelines criminal history score on the outcome in those districts as compared to non-border districts.

Defendants within individual districts are more likely to share sociodemographic characteristics than with defendants in other districts because of the tendency in at least some parts of the United States to be more heterogenic in their populations. Traditional regression models unfortunately tend to ignore these kinds of correlations between defendants sentenced in the same jurisdiction.

In addition, the theory of courtroom communities is relevant. Sentences of defendants in the same district may be more correlated because they share the same courtroom cultures and sentencing judges than they are correlated with sentences issued in other districts exhibiting different cultures and judges. These group-based factors, resulting from individuals nested in districts, may also impact sentencing outcomes.

The statistical issue, then, when criminal defendants are nested in a higher level, such as district courts in the federal context, is that assuming that penalty outcomes for the dependent variable are independent from the higher level may be erroneous.[180] In such a case, the single-level regression model's assumption of independence of outcomes may be violated, rendering results that may produce

---

[177] Peter C. Austin et al., *An Introduction to Multilevel Regression Models*, 92 Canadian J. Pub. Health 50, 50 (2001).

[178] U.S. Sent'g Comm'n, Illegal Reentry Cases 8 (2015).

[179] Michael T. Light, Michael Massoglia, & Ryan D. King, *Citizenship and Punishment: The Salience of National Membership in U.S. Criminal Courts*, 79 Am. Soc. Rev. 4 (Online Supp. 2014). Foreign convictions are not formally counted in the official Guidelines criminal history calculation but they may be considered for purposes of upwardly departing because the official calculation underestimates the true criminal background. U.S. Sentencing Guideline Manual § 4A1.2(h).

[180] Noelle E. Fearn, *A Multilevel Analysis of Community Effects on Criminal Sentencing*, 22 Just. Q. 452, 457 (2005).

biased estimates and misestimate standard errors.[181] Importantly, there is now available a sophisticated statistical procedure that can address these concerns when data is nested—multilevel modeling. In sum, "the utility of multilevel models lies in their capacity to aggregate cases by group membership and to test simultaneously for individual and group effects on the dependent variable."[182]

B. The Benefits of Multilevel Regression Models

Multilevel analyses, when suitable for the data, are able to provide numerous benefits over single-level regression models. First, multilevel methods can account for the lack of independence when individuals are nested in groups.[183] Multilevel modeling does not assume that the impact of an explanatory variable is the same across groups. Instead, multilevel models can be specified to account for between-group variability in explanatory variables and residuals.[184] Second, the methodology is preferable to simply controlling for the group-level effect as can be done in a single-level regression model. Multilevel modeling can simultaneously test the effects of both individual and group explanatory variables on the outcome of interest.[185] A multilevel model is able to indicate whether the individual-based explanatory factors impact the outcome variable while also indicating how group characteristics affect the relationships between the individual factors and the outcome of interest.[186]

Third, multilevel models are not limited to two levels; they can accommodate additional levels. As an illustration, multilevel regressions are popular in educational research where students are nested in classrooms which are nested in schools. The current challenge of including multiple levels is the substantial increase in computer resource capacity that is necessary to run a model with numerous explanatory factors included. An attractive feature is that there need not be the same number of units at each level. Nor must the levels be strictly hierarchical in nature. They may merely be nested. Thus, a multilevel model can be cross-level, such as defendants nested in years and nested in districts. Such a design would account, then, for both annual and regional variables.

Fourth, multilevel models partition the overall variance in the outcome of interest among the levels of analysis (e.g., at the individual level and then at the group level). The result indicates how much of the variation in the outcome is accounted for by the grouping.

---

[181] Austin et al., *supra* note 179, at 50. A violation of the assumption of independence can produce Type 1 errors. James L. Peugh, *A Practical Guide to Multilevel Modeling*, 48 J. School Psychol. 85, 86 (2010).

[182] Weidner et al., *supra* note 99, at 410.

[183] *Id*. (noting in single-level regressions the lack of independence may exaggerate the significance of the parameter estimate).

[184] Brian D. Johnson, *Cross-Classified Multilevel Models: An Application to the Criminal Case Processing of Indicted Terrorists*, 28 J. Quant. Criminology 163, 171 (2012).

[185] Fearn, *supra* note 182, at 468. In even more technical terms, "multilevel techniques take into account variance at both the individual and group levels, thus allowing intercepts and slope coefficients for selected variables to vary across groups." Stephen R. Porter & Paul D. Umbach, *Analyzing Faculty Workload Data Using Multilevel Modeling*, 42 Res. Higher Educ. 171, 177 (2001).

[186] Porter & Umbach, *supra* note 187, at 178.

C. Step One: Running the Null Model

The initial step in a multilevel model project is to run a null model. The null model is also referred to as an unconditional model because it has no explanatory factors included. The purpose is to statistically obtain the intraclass correlation coefficient ("ICC") to determine if multilevel modeling is appropriate for the data. The ICC provides the proportion of the total variance in the outcome that is accounted for by the clustering at the nested group level. In other words, for purposes of this study, the statistic is a measure of how much of the differences in upward departure decisions are attributable to variations in district court practices. If the ICC indicates that intraclass correlation exists with statistical significance, the assumption of independence required by the single-level regression model may be rejected and the data are appropriate for multilevel modeling.[187] Still, even if the ICC shows statistical significance, if it is not practically significant, the researcher can still reasonably decline to model that level. Multilevel analysis with numerous explanatory variables to test requires complex algorithmic processing. An ICC that provides a statistically significant, though practically small, proportional variance may convince the researcher that the ability to include more explanatory variables at the lower levels may outweigh any interest in retaining the practically unimportant variation at that nested level.[188]

D. Three-Level Null Models for the Upward Departure Dataset

Multilevel models, like single-level regression models, are commonly tested on continuous dependent variables. But when the outcome of interest is binary in nature, different modeling must be employed because a binary dependent variable means that the normal assumptions of a normally distributed response variable and homoscedatic errors are violated.[189] In the study presented herein, the outcome of interest is binary, being whether an upward departure was ordered (or not). Statistical techniques can be employed to transform such a binary outcome to achieve normality and reduce heteroscedasticity, typically through the logit function,[190] as was used herein.

A statistical model to fit data with a binary dependent variable is called a generalized linear model with three components: (1) a linear regression equation, (2) a specific error distribution, and (3) a nonlinear link function that transforms the predicted values for the dependent variable to the observed values.[191]

For the study herein, the binary response variable for the ith defendant in district j, is:

---

[187]   J. Kyle Roberts, *An Introductory Primer on Multilevel and Hierarchical Linear Models*, 2 Learning Disabilities 30, 32 (2004).

[188]   Tom A.B. Snijders, *Fixed and Random Effects, in* Encyclopedia of Statistics in Behavioral Science 664, 665 (Brian S. Everitt & David C. Howell eds., 2005). At times, there is a give-and-take between resource capabilities and theoretical interests.

[189]   Joop J. Hox & Cora J.M. Maas, *Multilevel Analysis, in* Encyclopedia of Social Measurement 785, 790 (Kimberly Kempf-Leonard ed., vol. 2, 2005).

[190]   *Id*.

[191]   *Id*.

$$Y_{ij} \begin{cases} 1 \text{ for upward departure} \\ 0 \text{ for no upward departure} \end{cases}$$

The transformation of the dichotomous dependent variable for an upward departure presented herein utilizes the logit link function.

$$\eta_{ij} = ln \left(\frac{p}{1-p} \right) \qquad \text{Logit Link Function}[192]$$

In the logit link function, the Greek letter eta (η) represents the transformed linear predictor. Exponentiating the resulting η parameter provides the odds ratio. The p is the probability of the outcome occurring and the denominator $(1 - p)$ is the probability of the outcome not occurring. The equation represents the odds of the outcome.

At the outset of this study, it was considered that a three-level model might be appropriate considering district courts are nested within the higher level circuit courts of appeal and/or within years, with the latter perhaps accounting for changes in sentencing patterns over time and using annual time periods as the temporal division.

A few statistical notes should be briefly mentioned before addressing the models. The software utilized for the study presented herein, including the three-level models that follow, was SPSS version 24. Further, there is no issue of selection bias and therefore no need for the so-called Heckman correction. Selection bias may occur when the researcher obtains data from a non-random sub-sample of the population of interest.[193] The relevant population of interest in this paper is federal defendants sentenced in the federal system during the period of study. The data analyses included herein were not limited to some sub-sample of that population.

In any event, the specification for a three-level null model is as follows:

$$\begin{aligned} \eta_{ij} &= \beta_{0jk} & \text{Level-1} \\ \beta_{0jk} &= \gamma_{00k} + \mu_{0jk} & \text{Level-2} \\ \gamma_{00k} &= \gamma_{000} + \mu_{00k} & \text{Level-3}[194] \end{aligned}$$

It was of interest, then, to test for whether the final model ought to account for serious nesting patterns which may introduce bias from the circuit courts of appeal as Level-3. The initial step in creating a multilevel model with three levels is to estimate the null model, which is provided in Table 5.

From Table 5 it is estimated that 7% of the variation in upward departures is between district courts and almost 2% of the variation is between circuit courts of appeal. The ICC was statistically significant for Level-2 district courts, yet was not significant for the Level-3 circuit courts. Practically, it was not surprising that there was not shown to be statistical significance with circuit courts. An earlier scan of bivariate data for the proportion of upward departures in the districts did not reveal consistencies for districts nested in circuits. Instead, the circuits tended

---

[192] RONALD H. HECK ET AL., MULTILEVEL MODELING OF CATEGORICAL OUTCOMES USING IBM SPSS 151 (2012).

[193] Shawn Bushway et al., *Is the Magic Still There? The Use of the Heckman Two-Step Correction for Selection Bias in Criminology*, 23 J. QUANTITATIVE CRIMINOLOGY 151, 152 (2007).

[194] HECK ET AL., *supra* note 194, at 183.

Table 5. Null Model for Upward Departures with Districts Nested in Circuits.

Fixed Effects	b	S.E.	
Intercept	-3.934	.087***	
Random Effects	s^2	S.E.	ρ
Level-1	3.29		
Level-2	.250	.042***	6.94%
Level-3	.060	.162	1.67%
-2LL=4324243			
n=623,947			

*** $p < .001$

to encompass a mix of low and high use of upward departures within their nested districts. For example, while three of the districts within the Fifth Circuit yielded the highest proportions of upward departures (Northern District of Texas at 6.5%, Western District of Louisiana at 5.7%, and Eastern District of Louisiana at 4.8%), the Fifth Circuit also included one district with a below-average rate of upward departures (Southern District of Texas at 1.5%). Overall, the Fifth Circuit ranked as the fifth highest among the 12 circuits in its total proportion of upward departures. The First Circuit ranked first overall, with a total of 3.3% of sentences with upward departures. But the First Circuit also presented with vastly different practices within its district court outcomes, as well. Most of the upward departures in the First Circuit were issued in the District of Puerto Rico (at 4.4%), yet this circuit also included the District of Rhode Island which issued one of the lowest rates of upward departures (at 0.5%).

While circuit court variation was not statistically significant, it alternatively was likely that there might be variations by time. Thus, a three-level null model was run for district courts nested in fiscal years, which is presented in Table 6.

Table 6. Null Model for Upward Departures for Districts Nested in Years.

Fixed Effects	b	S.E.	
Intercept	-3.937	.057***	
Random Effects	s^2	S.E.	ρ
Level-1	3.29		
Level-2	.282	.046***	7.69%
Level-3	.093	.010***	2.54%
-2LL: 4328082			
n=623,947			

*** $p < .001$

This null model with district courts nested in fiscal years demonstrated that 8% of the variation in upward departures is between district courts. It was also found that there is a statistically significant variation with Level-3 being an annual indicator. Yet, for several reasons, the nesting of upward departure outcomes at a level with years was dropped to proceed with a more developed two-level model. The ICC for years was, in practical terms, indicating a low degree of variation by year at

less than 3%. As multiple explanatory variables were expected to be included in the final model with both fixed and random effects, a three-level model including years would present as an extremely complicated model from a computing resource perspective. Indeed, as will be indicated below, even in a two-level design with district courts at the higher grouping, the final model had to be curtailed a bit because of convergence issues when attempting to model all independent variables as both fixed and random effects. An additional concern is that there were only 8 groups involved for years (i.e., eight consecutive fiscal years), an extremely low number for multilevel modeling purposes. In any event, as a primary interest for this study was regional variations in discretionary sentencing decisions, the Level-3 variation with years was dropped. Still, the three-level model indicated in Table 6 was presented herein for informational purposes.

E. The Two-Level Null Model for the Upward Departure Dataset

As the three-level designs just summarized were vetoed, a null model with two levels to account for nesting in districts could be run. The null model for two-level design with a dichotomous dependent is specified with the following equations.

$\eta_{ij} = \beta_{0j}$ Level-1 Null Model
$\beta_{0j} = \gamma_{00} + \mu_{0j}$ Level-2 Null Model

In these null models for this study, the term β_{0j} is the intercept, which is the average log odds of an upward departure in group j. At Level-2, the term γ_{00} represents the fixed intercept, being the log odds of an upward departure in a typical district for the average individual. The variance parameter μ_{0j} is the random intercept and signifies the variability of the outcome across Level-2 groups.[195]

In a generalized linear multilevel model using a logit link because of a binary response variable, the Level-1 residuals are assumed to follow the standard logistic distribution, with a mean of 0 and a variance ($\hat{\sigma}^2$) set to $\pi^2/3$, which is equal to 3.29. For a dichotomous outcome, the intraclass correlation coefficient (i.e., a statistic that indicates the proportion of total variability in outcomes which arises at the higher level) is computed in a two-level model as:

$$\frac{\tau_{00}}{\tau_{00} + 3.29}$$ Intraclass Correlation Coefficient (ICC)

The term τ_{00} represents the between-group variance at Level-2.[196]

Table 7 provides for the null model results for upward departures where Level-1 are individual defendants and Level-2 are district courts. Table 7 is the basis for the final model contained in Table 3 in the main body of this Article.

---

[195] Heck et al., *supra* note 194, at 151.
[196] *Id*. at 94.

Table 7. Null Model for Upward Departures Nested in Districts.

Fixed Effects Intercept	b -3.921	S.E. .058***	
Random Effects	s^2	S.E.	ρ
Level-1	3.29	---	
Level-2	.301	.047***	8.38%
-2LL=4324129 n=623,947			

*** *p* < .001

The ICC computed for the two-level null model means that 8% of the variability in upward departures is accounted for by districts.[197] This result is relatively within the bounds of other studies of federal sentencing. The other research that report on the partition of variance results typically find that between 4 and 12% of the variance in sentence length was accounted for at the districts level, with the exactly percentage depending on the period studied, the crimes included, and when reporting full models, the control variables used.[198]

As expected from the courtroom communities' perspective, the Level-2 random effect is significant at the .001 level, which indicates that the probability of an upward departure significantly varies between districts. Indeed, in a separate analysis to compare district means, wide variation in proportions were observed. The proportion of upward departures at the district court level ranges from a low of 0.5% (Northern District of Oklahoma, District of New Mexico, and District of Rhode Island) to a high of 6.5% (Northern District of Texas). Thus, the district with the greatest proportion of upward departures is more than twelve times that of the district with the lowest percentage, indicating a stark district level differential.

The intercept in the two-level null model represents an estimate that can be converted to the overall probability of an upward departure. The random effect represents the degree to which the outcome varies across federal districts. The estimated probability of a defendant receiving an upward departure in the average district is approximately 2%.[199]

Once the researcher chooses the null model with the appropriate higher level(s), the researcher can add explanatory factors. In a very simple model, we can

---

[197] This leaves 92% of the variability to be accounted for at the individual case level (or other unknown factors).

[198] Light, *Noncitizens*, *supra* note 120, at 462 (4% variance in length of sentence and 5% in sentences requiring incarceration); Lynch & Omori, *supra* note 104, at 429 (11% for drug trafficking crimes); Feldmeyer & Ulmer, *supra* note 39, at 250 (7%); Farrell et al., *supra* note 113, at 112 (5% for length of incarceration and 8% of the variance for the odds of incarceration was between districts); Albonetti & Baller, *supra* note 39, at 64 (12% for drug trafficking crimes); Kautt, *supra* note 90, at 653 (7% for drug trafficking crimes).

[199] The formula to obtain the overall expected proportion is an inverse of the logit link function: $[(1/(1 + e^{-\eta})) \times 100\%]$. Plugging in the coefficient for the fixed effect coefficient, the formula becomes $[(1/(1 + e^{3.921})) \times 100\%]$, which is equal to 1.94%.

add a Level-1 explanatory variable and a Level-2 predictor, such as the following equation illustrates.

$$\eta_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} \qquad \text{Level-1}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + \mu_{0j} \qquad \text{Level-2}$$
$$\beta_{1j} = \gamma_{10} + \mu_{1j}$$

Now γ_{00} is the log odds that the outcome = 1 when explanatory variable X = 0 and μ = 0. β_1 is the log odds effect that the outcome is = 1 for every one unit increase in the variable X in group *j*. To get a more interpretable result for the effect of X, we can exponentiate β_1 to obtain the odds ratio to compare the odds for individuals spaced one unit apart on X. Then W*j* represents the random effect of that predictor variable in group *j*.

In this study, the null model with district courts at Level-2 was the choice and the independent variables that survived into the final model are provided in Table 3 in the main body of the text. In Table 3, the ICC statistic indicates that 2% of the overall variance remains with district courts. The intraclass coefficient is no longer statistically significant when accounting for multiple fixed and random effects. Nonetheless, the substantial reduction in the -2 Log-Likelihood statistic between the null model and the full model indicates a significantly better fit of the full model for this dataset. Further discussion on methodological choices along the way to the final model is next.

F. Transforming Variables and Excluding Factors Regarding the Full Model

Some variables were transformed for the final model as explained below. In addition, other factors were tested yet eliminated in the end for the reasons ascribed to them herein.

For purposes of the descriptive statistics in Table 2, the variables for final offense level, criminal history, and number of counts are in their original metrics. For the multilevel model in Table 3, these three variables are each grand mean centered for ease of interpretation as none of them can have zero as a real value. In federal sentencing, defendants must have at least one count of conviction, the lowest criminal history category is I (i.e., 1), and the minimum offense severity level is 1. In a logistic model, the intercept is interpreted to mean the value of the outcome when all predictors are equal to 0. This has no practical meaning for variables that cannot actually have a real world value of 0, which is the case for these three variables. Grand mean centering is the statistical convention for adjusting the metrics to have a more interpretable intercept in such a case.

The number of counts (of conviction) variable was transformed for statistical purposes. In the original data, the number of counts variable was skewed to the right. This variable was first centered at the grand mean. Then to enable a natural log transformation to adjust for the skew and more closely approximate a normal distribution, the value of .1 was added to the mean centered variable because log transformations are not possible on values of 0.

Race/ethnicity was originally coded as dummy variables of black, Hispanic, and other, with white as the reference category. In a full multilevel model with such coding with all fixed effects, the only statistically significant result was for the category of other as compared to whites. This result is practically meaningless

because the grouping of "other" includes a heterogeneous mix of native Alaskan, native American, non-U.S. American Indians, Asian, Pacific Islander, multi-racial, and a smaller subset of other.[200] In addition, SPSS could not properly compute a random effect for this variable with this coding scheme involving three dummy variables. As race/ethnicity is such an important topic of interest in criminal justice, it seemed more worthwhile to recode the variable as a single dichotomous factor in order to incorporate a race-based variable in the formula and to be able to model it with both fixed and random effects.

The full model includes all 94 district courts. This is mentioned because many studies that incorporate district courts in their variables exclude the districts that are in the U.S. territories (Puerto Rico, Virgin Islands, Guam, North Mariana Islands). These researchers argue the territories are viewed as different because states enjoy greater rights than them and, thus, the inclusion of the territories may introduce nonrandom bias.[201] However, other experts challenge the assumption of substantive differences between districts courts within the states and those in the territories.[202] Indeed, researchers in at least one study found far more similarities than differences in sentencing outcomes, except that the districts in the territories tended to be more punitive.[203] These researchers further contend that excluding the territories actually may do more harm by not portraying an accurate picture of the salience of the Guidelines and judicial compliance with them from a national perspective.[204] I determined it was preferable to include the territories for similar reasons.

The general offense type was excluded from the random effects due to the complexity of the algorithm necessary to compute a multilevel model with them included. In other words, the model with the offense type having random effects was overly complicated for computational iterations, resulting in a failure of convergence. Convergence was achieved after excluding offense types at Level-2, while still retaining their Level-1 fixed effects.

It is noted that four additional independent variables were tested but removed before the final model for reasons of parsimony and specific statistical challenges. The applicability of a mandatory minimum statute was not statistically significant (at the .001 level) at Level-1 in any model and thus was removed as there was no theoretical justification to retain it as a factor in a study on *upward* departure outcomes. A variable tied to the Guidelines-recommended sentence was removed because of multicollinearity concerns with the final offense level and criminal history score variables. Notably, all independent variables attempted in any model were tested for multicollinearity. For the independent variables retained and shown in the final model in Table 3, results indicated no significant collinearity problems. All variance inflation factor scores resided within an acceptable level (VIFs < 3). A variable regarding the guideline recommended sentence had previously triggered multicollinearity concerns (with some VIFs greater than 5) and was therefore removed.

A series of dummy variables to distinguish fiscal years of sentencing were also

---

[200] U.S. Sent'g Comm'n, Variable Codebook for Individual Offenders 31 (2015).

[201] *E.g.*, Farrell et al., *supra* note 113, at 103 n. 75; Kautt, *supra* note 90, at 648. *See also* Light, *Noncitizens*, *supra* note 120, at 456 (excluding the territories without stating reason).

[202] Gail Iles et al., *U.S. Territorial Exclusion in Federal Sentencing Research: Can it be Justified?*, 3 Int'l J. Criminology & Soc. 113, 113 (2014).

[203] *Id.* at 122.

[204] *Id.* at 113.

dropped. While the annual rates of upward departures were statistically significant compared to 2008 as the dummy, the overall statistical impact (according to F statistic results) on explaining upward variances for the timing factor was among the weakest among the various explanatory variables. The statistical resources necessary to account for the seven dummy variables for years did not then seem worthwhile.

Another variable was tested and also dropped. No statistically significant effects of education level on upward departures were observed in any tested model. Without any pressing need to focus on educational level as it does not represent the most egregious type of discriminatory category, it was discarded as an explanatory factor.

As a final methodological note, the results here may advise other researchers that it might be preferable to model the main Guidelines proxies for crime severity and criminal background with the two separate factors of final offense level and final criminal history category, respectively, rather than their combination as indicated by the Guidelines' minimum sentence recommendation. As shown herein, the two variables may actually have the opposite effect on the outcome of interest, which would unfortunately be indiscernible when using the minimum sentence combination instead.